

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

Algumas Propostas para Imputação de
Dados Faltantes em Teoria de Resposta ao Item

por

Edna Alessandra Pereira

Orientador: Prof. Dr. Antônio Eduardo Gomes

Julho de 2014

Edna Alessandra Pereira

Algumas Propostas para Imputação de Dados Faltantes Em Teoria de Resposta ao Item

Dissertação apresentada ao Departamento de
Estatística do Instituto de Ciências Exatas
da Universidade de Brasília como requisito
parcial a obtenção do título de Mestre em
Estatística.

Universidade de Brasília

Brasília, Julho de 2014

TERMO DE APROVAÇÃO

Edna Alessandra Pereira

ALGUMAS PROPOSTAS PARA IMPUTAÇÃO DE
DADOS FALTANTES EM TEORIA DE RESPOSTA AO ITEM

Dissertação apresentada ao Departamento de
Estatística do Instituto de Ciências Exatas
da Universidade de Brasília como requisito
parcial a obtenção do título de Mestre em
Estatística.

Data da defesa: 07 de Julho de 2014

Orientador: Prof. Dr. Antônio Eduardo Gomes

Universidade de Brasília
Brasília, Julho de 2014

*à Deus, às minhas filhas, aos meus pais,
aos meus irmãos e toda minha família
aos amigos, aos meus chefes e colegas de trabalho*

Agradecimentos

Agradeço a Deus e a Nossa Senhora, primeiramente, por terem me acompanhado e me iluminado em todas as horas. Por me capacitar sempre!

As minha filhas Lis, Nicole, Anita e Giovana. Por elas que eu luto. Minha vida é delas. Minha vida é elas!

Uma dedicatória especial para o meu falecido pai, Edson Pereira da Silva, pelo amor, pelos ensinamentos e por sempre estar torcendo por mim. Pelo amor, pelo carinho e afeto.

A minha querida mãe, Celica Maria Pereira da Silva, aos meus irmãos e todos os meus familiares que torcem por mim.

Ao meu orientador, por acreditar em mim, por me mostrar os caminhos e por ser um exemplo de profissional.

Aos professores membros da banca examinadora: Fernando A. S. Moura, Eduardo Y. Nakano e Cibele Q. da Silva.

A todos os funcionários e professores da Universidade de Brasília.

Aos meus chefes Laura Bernardes, Eduardo Moreno e Maria das Graças, pela confiança e pela força.

E finalmente aos meus colegas de trabalho que tanto torcem por mim.

Sumário

Agradecimentos	ii
Lista de Figuras	9
Resumo	10
Abstract	11
Motivação	12
Revisão da Literatura	15
1 Conceitos Básicos de Teoria de Resposta ao Item	17
1.1 Introdução	17
1.1.1 Teoria Clássica dos Testes e a Teoria de Resposta ao Item . .	18
1.1.2 Pressupostos da Teoria de Resposta ao Item	23
1.2 Modelos Unidimensionais da Teoria de Resposta ao Item	24
1.2.1 Modelos Dicotômicos	25
1.2.2 Modelos Não Dicotômicos	31
1.2.3 Métodos de Estimação dos Parâmetros	38
1.2.4 Teoria de Resposta ao Item Multidimensional	40
1.2.5 Modelos Multidimensionais da Teoria de Resposta ao Item . .	44
2 Dados Faltantes	46
2.1 Introdução	46
2.1.1 Mecanismos de Dados Faltantes	48
2.1.2 Métodos de Imputação de dados faltantes	49

2.1.3	Dados Faltantes em Teoria de Resposta ao Item	56
3	Estudos de Simulação	59
3.1	Introdução	59
3.1.1	Metodologia para simulação	60
3.1.2	Caso 1: Simulação sem dados faltantes	62
3.1.3	Caso 2: Simulação utilizando dados faltantes ignorados	65
3.1.4	Caso 3: Simulação com dados faltantes como respostas incorretas	67
3.1.5	Caso 4: Simulação ignorando dados faltantes no processo de calibração e estimativas das proficiências considerando os dados faltantes como respostas incorretas.	69
4	Propostas para probabilidade de imputação	74
4.1	Estimação sem imputação	76
4.1.1	Estimação sem imputação para os valores da proficiência	76
4.1.2	Estimação sem imputação para os valores do parâmetro de di- ficuldade dos itens	78
4.1.3	Estimação sem imputação para os valores do parâmetro de dis- criminação dos itens	79
4.2	Exemplo de proposta para probabilidade de imputação utilizando re- gressão logística	80
4.2.1	Estimação com imputação via regressão logística para os valores da proficiência	83
4.2.2	Estimação com imputação via regressão logística para os valores dos parâmetros de dificuldade dos itens	84
4.2.3	Estimação com imputação via regressão logística para os valores dos parâmetros de discriminação dos itens	86
4.2.4	Estimação com imputação via regressão logística para os valores dos parâmetros de acerto ao acaso	88
4.3	Regressão isotônica	89
4.3.1	Estimação com imputação via regressão isotônica suavizada para os valores dos parâmetros da proficiência	91

4.3.2	Estimação com imputação via regressão isotônica suavizada para os valores dos parâmetros de dificuldade dos itens	92
4.3.3	Estimação com imputação via regressão isotônica suavizada para os valores dos parâmetros de discriminação dos itens. . .	94
4.3.4	Estimação com imputação via regressão isotônica suavizada para os valores dos parâmetros de acerto ao acaso.	96
4.4	Comparação entre os métodos de estimação das proficiências: sem imputação, com imputação via regressão logística e com imputação via regressão isotônica	98
5	Conclusões e Trabalhos Futuros	101
5.1	Conclusões	101
5.2	Trabalhos Futuros	102
	Referências Bibliográficas	103
5.2.1	Regressão Logística	110

Lista de Figuras

1.1	Curva Característica do Item: [a] item 1: $b=0$ [b] item 1: $b=-1$, item 2: $b=0$ e item 3: $b=1$	26
1.2	Curva Característica do Item : [a] item 1: $a = 0$; 2 e $b = 0$, item 2: $a = 1$ e $b = 0$, item 3: $a = 3$ e $b = 0$ [b] item 1: $a = 1$ e $b = 1$, item 2: $a = 3$ e $b = -1$, item 3: $a = 3$ e $b = 1$, item 4: $a = 1$ e $b = -1$	29
1.3	Curva Característica do Item: [a] item 1: $a = 1$ e $b=0$, item 2: $a = 1$ e $b = 0$, item 3: $a = 3$ e $b = 0$ [b] item 1: $a = 1$, $b = 1$ e $c = 0:3$; item 2: $a = 3$, $b = -1$ e $c = 0:1$; item 3: $a = 3$ e $b = 1$ e $c = 0:1$; item 4: $a = 1$, $b = -1$ e $c = 0:3$	30
1.4	Curva Característica de dois Itens para o Modelo Nominal	32
1.5	Curva Característica do Item para o Modelo de Escala Gradual	34
3.1	Caso 1: Simulação sem dados faltantes. Diagrama de dispersão entre o parâmetro de discriminação do item gerado (valores reais dos parâmetros) e o parâmetro de discriminação do item estimado	62
3.2	Caso 1: Simulação sem dados faltantes. Diagrama de dispersão entre o parâmetro de dificuldade do item gerado (valores reais dos parâmetros) e o parâmetro de dificuldade do item estimado	63
3.3	Caso 1: Simulação sem dados faltantes. Diagrama de dispersão entre a proficiência gerada e a proficiência estimada	64
3.4	Caso 2: Simulação com dados faltantes ignorados. Diagrama de dispersão entre o parâmetro de discriminação do item gerado (valores reais dos parâmetros) e o parâmetro de discriminação do item estimado	65

3.5	Caso 2: Simulação com dados faltantes ignorados. Diagrama de dispersão entre o parâmetro de dificuldade do item gerado (valores reais dos parâmetros) e o parâmetro de dificuldade do item estimado . . .	66
3.6	Caso 2: Simulação com dados faltantes ignorados. Diagrama de dispersão entre a proficiência gerada (valores reais dos parâmetros) e a proficiência estimada	67
3.7	Caso 3: Simulação com dados faltantes como respostas incorretas. Diagrama de dispersão entre o parâmetro de discriminação do item gerado (valores reais dos parâmetros) e o parâmetro de discriminação do item estimado	68
3.8	Caso 3: Simulação com dados faltantes como respostas incorretas. Diagrama de dispersão entre o parâmetro de dificuldade do item gerado e o parâmetro de dificuldade do item estimado	69
3.9	Caso 3: Simulação com dados faltantes como respostas incorretas. Diagrama de dispersão entre a proficiência gerada (valores reais dos parâmetros) e a proficiência estimada	70
3.10	Caso 4: Simulação com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências. Diagrama de dispersão entre o parâmetro de discriminação do item gerado (valores reais dos parâmetros) e o parâmetro de discriminação do item estimado	71
3.11	Caso 4: Simulação com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências. Diagrama de dispersão entre o parâmetro de dificuldade do item gerado (valores reais dos parâmetros) e o parâmetro de dificuldade do item estimado	72
3.12	Caso 4: Simulação com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências. Diagrama de dispersão entre a proficiência gerada (valores reais dos parâmetros) e a proficiência estimada	73

4.1	[a] Diagrama de dispersão entre a proficiência gerada (valores reais do parâmetro) e a proficiência estimada sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências). [b] diagrama de dispersão do erro entre os valores da proficiência gerada e a proficiência estimada.	77
4.2	Diagrama de dispersão entre a proficiência estimada, através da matriz completa de dados, e a proficiência estimada sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências). .	77
4.3	Diagrama de dispersão entre a dificuldade do item gerado (valores reais do parâmetro) e a dificuldade do item estimado sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).	78
4.4	Diagrama de dispersão entre os valores das estimativas dos parâmetros de dificuldade do item estimado através da matriz completa de dados, e as estimativas do parâmetro de dificuldade do item sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).	79
4.5	Diagrama de dispersão entre a discriminação do item gerado (valores reais do parâmetro) e a discriminação do item estimado sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).	80
4.6	Diagrama de dispersão entre a discriminação do item estimado através da matriz completa de dados, e a discriminação do item estimado sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).	80
4.7	Exemplos de ajuste da curva logística para cálculo das probabilidades de imputação.	81

4.8	[a] Diagrama de dispersão entre os valores estimados da proficiência via regressão logística e a proficiência gerada (valores reais do parâmetro)	
	[b] diagrama de dispersão do erro entre os valores estimados da proficiência via regressão logística e a proficiência gerada (valores reais do parâmetro).	83
4.9	Diagrama de dispersão entre as estimativas da proficiência com imputação via regressão logística e os valores das estimativas do parâmetro da proficiência estimado através da matriz completa de dados.	84
4.10	Diagrama de dispersão entre as estimativas do parâmetro de dificuldade com imputação via regressão logística e os valores das estimativas do parâmetro de dificuldade gerado (valores reais do parâmetro.	85
4.11	Diagrama de dispersão entre as estimativas do parâmetro de dificuldade com imputação via regressão logística e os valores das estimativas do parâmetro de dificuldade sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).	85
4.12	Diagrama de dispersão entre as estimativas do parâmetro de dificuldade com imputação via regressão logística e os valores das estimativas do parâmetro de dificuldade estimado através da matriz completa de dados.	86
4.13	Diagrama de dispersão entre as estimativas do parâmetro de discriminação com imputação via regressão logística e o parâmetro de discriminação gerado (valores reais do parâmetro).	86
4.14	Diagrama de dispersão entre as estimativas do parâmetro de discriminação com imputação via regressão logística e os valores das estimativas do parâmetro de discriminação sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).	87
4.15	Diagrama de dispersão entre as estimativas do parâmetro de discriminação com imputação via regressão logística e os valores das estimativas do parâmetro de discriminação estimado através da matriz completa de dados.	87

4.16	Diagrama de dispersão entre as estimativas do parâmetro de acerto ao acaso com imputação via regressão logística e os valores das estimativas do parâmetro de acerto ao acaso sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).	88
4.17	Diagrama de dispersão entre as estimativas do parâmetro de acerto ao acaso com imputação via regressão logística e os valores das estimativas do parâmetro de acerto ao acaso estimado através da matriz completa de dados.	89
4.18	Exemplos de ajuste da curva isotônica para cálculo das probabilidades de imputação	90
4.19	[a] Diagrama de dispersão entre os valores estimados da proficiência via regressão isotônica suavizada e a proficiência gerada (valores reais do parâmetro) [b] diagrama de dispersão do erro entre os valores os valores estimados da proficiência via regressão isotônica suavizada e a proficiência gerada (valores reais do parâmetro).	91
4.20	Diagrama de dispersão entre as estimativas da proficiência com imputação via regressão isotônica suavizada e os valores das estimativas de proficiência estimado através da matriz completa de dados.	92
4.21	Diagrama de dispersão entre as estimativas do parâmetro de dificuldade com imputação via regressão isotônica suavizada e a dificuldade do item gerado (valores reais do parâmetro).	93
4.22	Diagrama de dispersão entre as estimativas do parâmetro de dificuldade com imputação via regressão isotônica suavizada e os valores das estimativas do parâmetro de dificuldade sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).	93
4.23	Diagrama de dispersão entre as estimativas do parâmetro de dificuldade com imputação via regressão isotônica suavizada e os valores das estimativas do parâmetro de dificuldade estimado através da matriz completa de dados.	94

4.24	Diagrama de dispersão entre as estimativas do parâmetro de discriminação com imputação via regressão isotônica suavizada e discriminação do item gerado (valores reais do parâmetro).	94
4.25	Diagrama de dispersão entre as estimativas do parâmetro de discriminação com imputação via regressão isotônica suavizada e os valores das estimativas do parâmetro de discriminação sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).	95
4.26	Diagrama de dispersão entre as estimativas do parâmetro de discriminação com imputação via regressão isotônica suavizada e os valores das estimativas do parâmetro de discriminação estimado através da matriz completa de dados.	96
4.27	Diagrama de dispersão entre as estimativas do parâmetro de acerto ao acaso com imputação via regressão isotônica suavizada e os valores das estimativas do parâmetro de acerto ao acaso sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).	96
4.28	Diagrama de dispersão entre as estimativas do parâmetro de acerto ao acaso com imputação via regressão isotônica suavizada e os valores das estimativas do parâmetro de acerto ao acaso estimado através da matriz completa de dados.	97
4.29	Regressão via kernel: [a] para o erro [b] desvio quadrático.	98
4.30	Regressão não paramétrica via kernel para o erro.	99
4.31	Regressão não paramétrica via kernel para o desvio quadrático.	100

Resumo

Neste trabalho, são estudadas duas propostas de métodos de imputação de dados para o ajuste do modelo de três parâmetros em estudos de teoria de resposta ao item. O primeiro método proposto consiste em utilizar como probabilidade de resposta positiva para imputação de resposta o valor fornecido pela curva de regressão logística das respostas certas e erradas do indivíduo como função das estimativas do parâmetro de dificuldade para os itens respondidos pelo indivíduo. Obtida a curva logística, a probabilidade de resposta positiva para um item que o indivíduo não tenha respondido é obtida pela curva logística ajustada calculada no valor estimado do parâmetro de dificuldade do item não respondido. A partir desta probabilidade, imputa-se uma resposta correta ou não. O segundo método é similar ao primeiro, mas utiliza para imputação a probabilidade de resposta correta dada pelo valor da curva suavizada via kernel fornecida pela regressão isotônica das respostas corretas e incorretas (1 e 0) como função decrescente das estimativas do parâmetro de dificuldade do itens respondidos. Num estudo de simulação, as estimativas dos parâmetros dos itens e da proficiência obtidas a partir dos dados imputados pelos dois métodos e também sem imputação de dados são comparadas com os valores reais e com aqueles obtidos para os dados completos.

Palavras Chave: *Palavras-chave: teoria de resposta ao item, dados faltantes, imputação, regressão logística, regressão isotônica, kernel..*

Abstract

In this work we have studied two imputation procedures for missing data when fitting the three parameters model in item response theory studies. The first proposed method uses as the probability of imputation of correct response the value obtained from the logistic regression of the correct (1) and incorrect (0) answers as a function of the difficulty parameter estimates for the items with responses. We take the fitted logistic curve and calculate the probability of imputation of positive response as a function of the value of the difficulty parameter estimates for the non responded items. The second method is similar to the first one, but we use the probability of imputation of a positive response provided by the kernel smoothed isotonic regression obtained taking the probability of positive response as a decreasing function of the estimated difficulty parameter for each item with a response. In a simulation study, the estimates of the item parameters and proficiency were compared to the real values and also to the estimates obtained for the data with no missing responses.

key words: *Keywords: item response theory, missing data, imputation, logistic regression, isotonic regression, kernel..*

Motivação

Os sistemas de bancos de dados surgiram na década de 60 com o objetivo de fornecer recursos capazes de armazenar, organizar, manipular e recuperar dados. Atualmente, existe um grande armazenamento das informações em bancos de dados de domínio público. Na área da educação, vários indicadores educacionais foram colocados à disposição dos especialistas, promotores de políticas públicas e do público em geral. Contudo, existe a necessidade de que esses indicadores sejam capazes de sintetizar as informações de forma eficaz e assim surgem, a cada dia, pesquisas e técnicas que objetivam resumir as informações.

Avaliações educacionais em grande escala ganharam grande importância em todas as esferas de governo, que passaram a usar cada vez mais seus resultados com o objetivo de melhorar seus investimentos e obter melhores resultados de rendimento escolar, buscando a qualidade da educação.

Ao aplicar uma avaliação para testar as habilidades dos alunos, e compará-las, um estudo essencial envolve as diversas interações entre os alunos e os itens (questões) do questionário. Procedimentos iniciais para análises de dados das avaliações educacionais foram baseados em métodos discutíveis, como a contagem do número de respostas corretas em um conjunto de itens de um teste. A suposição era de que pessoas com maior número de acertos tinham maior habilidade que aqueles que tinham menos respostas corretas. Responder uma questão corretamente era equivalente a responder qualquer outra questão corretamente. Porém, analistas de testes entenderam que alguns itens são mais difíceis do que outros.

Medidas de poder de discriminação do item, proporção de pessoas que escolheram cada alternativa e outros indicadores estatísticos foram então coletados para descrever funções de itens de um teste. Métodos de teoria de resposta ao item (unidi-

mensionais) também foram desenvolvidos. Estes métodos descrevem o funcionamento dos itens do teste para as pessoas em diferentes níveis de uma classificação contínua de habilidade ou conhecimento.

Contudo, um dos grandes problemas recorrentes nos bancos de dados que armazenam os resultados das avaliações são os dados faltantes (missing). Geralmente, os dados faltantes podem surgir tanto pela falta do aluno ao teste - nesse caso, nenhum quesito desse aluno é respondido - quanto pela falta de resposta a um ou mais itens do questionário respondido pelo aluno que respondeu o teste.

O problema da não resposta pode estar relacionado a uma série de fatores: alunos faltam, são avisados para ficarem em casa, professores não levarem muito a sério, pois a avaliação não é para nota, falta de tempo para realizar a prova, falta de motivação, entre outros. Um caso especial de dados faltantes é citado por Rose, von Davier e Wu (2010) onde, em uma avaliação de desempenho internacional, o problema da não resposta de alguns itens do aluno pode estar ligado às diferentes tradições de cada país. Nesse caso, a tendência de omitir respostas não é independente.

Como muitas técnicas estatísticas foram feitas para analisar dados completos, é frequente ignorar os alunos que não responderam todos os itens. Se esses respondentes retirados da análise possuem características diferentes dos demais, a inferência fica prejudicada. Por isso, métodos de imputação de dados foram desenvolvidos para lidar com o problema. Por estes métodos os dados faltantes são preenchidos com valores plausíveis, de acordo com a análise do perfil dos respondentes e diante das respostas aos itens. Estes dados faltantes geralmente são tratados ou como observações ausentes (ou seja, os dados faltantes são ignorados e as análises são feitas utilizando apenas os dados preenchidos) ou os dados faltantes são considerados como respostas erradas (como se toda a prova tivesse sido respondida e como se o aluno não tivesse habilidade suficiente para responder ao quesito que deixou em branco).

Segundo Rubin (Rubin, 1976; Rubin e Little, 2002) existem três tipos de dados faltantes: completamente aleatório (MCAR), quando o dado faltante não está relacionado a nenhuma resposta dos sujeitos; aleatório (MAR), quando o dado perdido é previsível através de outras variáveis do banco de dados; e não aleatório (MNAR), quando o dado faltante está relacionado com valores não observados. Dados MCAR e MAR não comprometem as inferências e podem ser considerados como ignoráveis. No

entanto, dados MNAR implicam na violação do princípio da ignorabilidade e exigem medidas adequadas para dar conta de seus efeitos na estimação dos parâmetros dos itens e nas proficiências dos respondentes.

Esta dissertação tem por objetivo estudar métodos de correção do efeito da não resposta na estimação dos parâmetros dos itens e da proficiência (habilidade) dos respondentes, para dados faltantes não ignoráveis. Simulações serão conduzidas para demonstrar os efeitos de diferentes tratamentos de dados faltantes. As omissões que são relacionadas às características dos respondentes são o alvo do presente estudo. Serão observados os efeitos de diferentes tratamentos de dados faltantes não ignoráveis, onde os dados faltantes são considerados como respostas incorretas, são ignorados ou são tratados em duas etapas: ignorados na fase de estimação dos parâmetros dos itens e tratados como resposta incorreta na fase da estimação da proficiência. Para imputação dos dados, consideraremos dois modelos: modelo de regressão logística e modelo de regressão isotônica suavizada.

Revisão da Literatura

Existem vários métodos para lidar com dados faltantes. No entanto, a maioria foi desenvolvida fora do contexto da Teoria de Resposta ao Item (TRI). Alguns métodos foram aplicados e analisados utilizando modelos tradicionais da TRI. Finch (2008) comparou o desempenho de várias técnicas de imputação, utilizando um conjunto de dados simulados em um modelo logístico de três parâmetros. Ele observou que vários destes métodos exibiram diferentes graus de eficiência. Sua investigação se limitou a escores ausentes para os mesmos quatro itens de um conjunto de 20 itens. Sheng e Carrière (2005) também analisaram as implicações de estratégias para tratar dados faltantes, utilizando o modelo de Rasch para dados de resposta ao item sob vários mecanismos de dados faltantes. A técnica bootstrap foi utilizada para imputação e observou-se que, para o caso de 20% de dados faltantes, este método parece ser a melhor estratégia para produção eficiente de estimadores consistentes e suas variâncias. O método bootstrap, no entanto, não funciona bem quando é grande a proporção dos dados faltantes, mesmo quando se tem um grande número de itens a serem explorados. Além disso, Sheng e Carrière limitaram suas investigações em medições contínuas. Huisman (2000) investigou os efeitos de vários métodos determinísticos de imputação naïve sobre as estimativas de habilidade latente dos entrevistados e utilizou o alfa de Cronbach como medida de confiabilidade do teste. Métodos determinísticos, no entanto, assumem que todas as condições que levam a escores dos itens ausentes são perfeitamente conhecidos, uma afirmação que nem sempre é válida. Sijtsma e van der Ark (2003) discutiram alguns métodos simples e propuseram dois métodos de imputação única não paramétricos, um dos quais parecia ser superior em recuperar várias propriedades estatísticas dos dados completos originais de um conjunto de dados incompletos. Além disso, van Ginkel, van der Ark, e Sijtsma (2007) mostraram

que versões de imputação múltipla de alguns métodos discutidos por Sijtsma e van der Ark produziram pouca discrepância em relação às propriedades estatísticas dos dados completamente observados. No entanto, as proporções de escores dos itens ausentes estudadas foram pequenas, entre 1% e 15%.

Capítulo 1

Conceitos Básicos de Teoria de Resposta ao Item

1.1 Introdução

A teoria de resposta ao item (TRI) surgiu na década de 50 substituindo a teoria clássica dos testes (TCT) como uma prática mais eficiente para medir características de indivíduos que não podem ser medidas diretamente. Ou seja, após a aplicação de um teste (ou pesquisa) a um conjunto de pessoas, a TRI busca um caminho para descrever as interações entre essas pessoas e os itens desse teste, considerando que as pessoas variam de acordo com suas características. Estas características são comumente denominadas de traço latente ou construto e podem ser medidas através de instrumentos de pesquisa, como questionários de desempenho para medir conhecimento e inteligência (por exemplo, medir a proficiência em matemática) e também questionários de atitude para medir personalidade e valores (por exemplo, medir o nível de qualidade de vida).

O uso da TRI em avaliações educacionais teve início no Brasil com o Sistema de Avaliação da Educação Básica (Saeb), em 1995, e, posteriormente, foi implementado no Exame Nacional para Certificação de Competências de Jovens e Adultos (Encceja), na Prova Brasil e, por último, no ENEM. No âmbito internacional, a TRI vem sendo utilizada largamente por diversos países: Estados Unidos, França, Holanda, Coréia do Sul, China e países participantes do Programa Internacional

de Avaliação de Estudantes (Pisa). Um dos grandes exemplos de avaliação que utiliza a TRI é o exame de proficiência em língua inglesa (TOEFL) (trecho Disponível em: <http://download.inep.gov.br/educacaobasica/enem/manuais/2012/guia-do-participante?Notas.pdf>. Acesso em : 09jun.2013).

1.1.1 Teoria Clássica dos Testes e a Teoria de Resposta ao Item

Na TCT, após a definição do traço latente (ou construto), é feita uma escolha sensata de itens e de escalas que devem representar bem o construto a ser medido. O construto medido pelos itens pode ser uma proficiência acadêmica ou de aptidão, ou pode ser uma atitude ou crença. As escalas dos respondentes aos itens podem ser dicotômicas (duas categorias), como certo ou errado, sim ou não, concordar ou discordar. Ou podem ser politômicas (mais de duas categorias) como um conjunto ordenado de respostas, numa escala de resposta tipo Likert, por exemplo. Na teoria clássica, a submissão de testes piloto e o cálculo das qualidades psicométricas do instrumento de pesquisa (validade, fidedignidade, α de Cronbach, etc.) servem para validar o instrumento de pesquisa e eliminar itens ruins (por exemplo, itens muito simples ou muito complicados, pois esses itens não diferenciam um respondente do outro).

Após a construção do teste é realizada a padronização dos escores brutos. Os escores brutos são as somas das respostas codificadas. Para itens dicotômicos as respostas podem ser codificadas em 0 e 1, sendo 0 a resposta errada e 1 a resposta correta. Para itens politômicos, as respostas podem ser codificadas, por exemplo, de 1 a 5 numa escala likert de 5 pontos. Ao somar todas as respostas dos itens do teste, os respondentes podem ser classificados de acordo com a soma total. Maiores somas significam maiores proficiências (ou maiores níveis de qualidade de vida), por exemplo.

O escore padronizado é um número que normaliza o escore bruto de cada respondente em função do resultado da aplicação do teste (ou pesquisa) para o conjunto de respondentes que o realizaram. Por isso mesmo, esta normalização é efetuada com base na média e o desvio padrão encontrados. Por exemplo, todo o respondente que

obtiver um número de acertos (escore bruto) igual a média do teste (ou pesquisa), receberá um escore padronizado de 500 pontos. A partir daí, a cada desvio padrão, acima ou abaixo da média, serão acrescidos ou retirados 100 pontos.

A análise clássica dos itens baseia-se em seus parâmetros descritivos, os quais auxiliam na interpretação da distribuição das respostas para cada alternativa. As características dos itens de um teste correspondem aos seguintes parâmetros: índice de dificuldade (proporção de participantes que responderam ao item corretamente); índice de discriminação, que mede a capacidade do item de diferenciar os participantes de maior habilidade (27 % dos respondentes com pontuações mais altas) daqueles de menor habilidade (27 % dos respondentes com pontuações mais baixas), correspondendo ‘a diferença entre a proporção de acertos do primeiro grupo e a do segundo grupo; e correlação bisserial entre a resposta numa dada categoria do item e o escore total do teste.

As principais estatísticas calculadas na TCT são:

- Escore total (respondente) = número de acertos. Varia de 0 a I (I=número de acertos);
- índice de dificuldade (item) = proporção ou porcentagem de acerto. Varia de 0 a 1;
- índice de discriminação (item) = diferença entre a porcentagem de acerto do grupo superior (27 % dos respondentes com os escores mais altos) e a porcentagem de acerto do grupo inferior (27 % dos respondentes com os escores mais baixos). Varia de 0 a 1;
- Coeficiente de correlação ponto bisseral (item) (r_{pbiss}): coeficiente de correlação de Pearson entre o escore total e a resposta ao item (1 = acerto, 0 = erro) de cada um dos respondentes. Em outras palavras, é uma medida de associação entre o desempenho no item e o desempenho na prova - estima a correlação entre a variável de desempenho no teste e uma variável latente (não observável) com distribuição normal que, por hipótese, representa a habilidade que determina o acerto ou erro do item. Varia de -1 a 1. A transformação do coeficiente de correlação ponto bisserial varia de 0 a 1.

A fórmula do coeficiente de correlação bisserial (item) é:

$$r_{biss} = \frac{M^+ - M^-}{S} \frac{p(1-p)}{h(p)} \quad (1.1)$$

Sendo M^+ a média da medida de desempenho para os alunos que acertaram o item, M^- a média da medida de desempenho no teste para os alunos que erraram o item, S o desvio-padrão da medida de desempenho no teste para todos os alunos, p o percentual de respostas e $h(p)$ o valor da densidade da distribuição normal com média 0 e variância 1 no ponto em que a área da curva à esquerda deste ponto é igual a p .

- Coeficiente α (teste): medida de fidedignidade do teste: varia de 0 a 1.

$$\alpha = \frac{I}{I-1} \left[1 - \frac{\sum_{n=i}^I p_i(1-p_i)}{s^2} \right] \quad (1.2)$$

Onde I é o número de itens, p_i é a proporção de acertos ao item i e s^2 é a variância dos escores.

- Erro-padrão de medida (teste) = medida de precisão do teste : valor maior ou igual a zero.

$$EPM = s\sqrt{1-\alpha} \quad (1.3)$$

- Assim, os dados da TCT são ajustados através do modelo:

$$Y_{jt} = \tau_j + \epsilon_{jt} \quad (1.4)$$

Onde:

Y_{jt} é escore do respondente j no teste t .

τ_j é o escore verdadeiro do respondente j (conhecimento).

ϵ_j é o componente de erro para o respondente j no teste t .

Suposição usual:

A TCT, porém, cria problemas importantes para as análises das interações entre as características das pessoas e as dos itens, tais como: os resultados encontrados dependem de um particular conjunto de itens (questões) que compõem o teste, ou seja, os resultados finais que medem um determinado construto estão associados ao teste aplicado como um todo (mudando o teste, mudam-se os resultados); a definição

do nível de dificuldade dos itens - por exemplo, em itens mais fáceis, médios ou difíceis - depende exclusivamente da “qualidade da amostra coletada” dos respondentes, não da qualidade “intrínseca” dos itens; a escolha do itens com melhor discriminação entre os respondentes é muito questionável já do ponto de vista conceitual; forçar as respostas de todos os respondentes a seguir uma distribuição perfeitamente gaussiana desnaturaliza o sentido da variância e do escore; etc.

A TRI é um conjunto de modelos matemáticos onde a probabilidade de um indivíduo acertar um item de um teste está expressa em função dos parâmetros do item e da habilidade do respondente. Em outras palavras, os modelos da teoria da resposta ao item mostram o relacionamento entre a habilidade ou o traço latente do indivíduo (simbolizado por θ) e a probabilidade de uma resposta positiva ao item de um teste.

Enquanto a TCT utiliza o escore do teste como referência de medida - tornando inviável a comparação entre indivíduos que não foram submetidos ao mesmo teste (ou pesquisa) - a TRI tem como foco principal o item e não o teste como um todo. O escore da TRI é frequentemente chamado de habilidade, traço ou proficiência e leva em conta o grau de dificuldade do item, a discriminação (capacidade do item em distinguir o nível de proficiência do aluno) e a probabilidade de acerto ao acaso (chute). A TRI foi desenvolvida para resolver o problema da TCT em relação a dependência que a medida da proficiência tem em relação ao teste aplicado e a dependência que os parâmetros do item têm em relação ao conjunto de respondentes. Em outras palavras, um item deve medir um determinado conhecimento e isso não pode depender de quem o está respondendo. Assim como a proficiência do respondente não pode depender dos itens que estão sendo apresentados a ele.

A grande importância de se usar TRI está em colocar todos os alunos na mesma métrica e a comparabilidade entre diferentes testes:

“A TRI possibilita a comparabilidade entre provas diferentes. Existem instrumentos para medir peso, altura, distância. Mas não há um instrumento que meça, de forma direta, o conhecimento. Então, para medir o conhecimento de uma pessoa há que fazê-lo de forma indireta e essa forma é a avaliação. Com a TRI, o que se pretende é criar uma unidade de medida para o conhecimento. Se você medir a altura de uma pessoa com uma trena ou com uma fita métrica, ela terá a mesma altura. Assim deve ser com o conhecimento: qualquer que seja a prova, uma vez que o participante carrega o mesmo conhecimento, ele deve obter a mesma nota.(Disponível em:

Resumindo, a TRI, através de um número grande de questões (itens) respondidas por uma amostra de indivíduos, obtém valores fixos para os parâmetros do item (probabilidade de acerto ao acaso, dificuldade e discriminação do item), através de técnicas de estimação (Máxima verossimilhança marginal, estimação bayesiana, etc), montando assim um banco de itens (questões). A partir do conhecimento dos valores desses parâmetros, as questões são posicionadas numa escala, permitindo, dessa forma, sua interpretação. O posicionamento de cada questão nos níveis da escala se dá a partir de critérios probabilísticos, os quais garantem que somente participantes com proficiência igual ou maior que a daquele nível possuem alta probabilidade de responder corretamente as questões que estão nesse nível e em níveis inferiores. Itens que conseguem discriminar entre os respondentes com maiores habilidades e aqueles com menores habilidades são muito mais confiáveis. Assim, os escores da TRI podem ter maior confiabilidade do que os escores dos itens utilizados na Teoria Clássica. Se diferentes examinados fazem diferentes testes, os escores da TRI são ajustados para as diferentes dificuldades. Os itens do teste são selecionados para corresponder a cada proficiência dos examinados, de modo que os examinados não fiquem entediados com itens fáceis, ou frustrados com itens muito difíceis. A TRI coloca os escores de diferentes testes na mesma métrica, de modo que cada examinado pode ter um teste personalizado. O Conhecimento da dificuldade do item é usado na construção dos testes para igualar os níveis das características da população. Por exemplo, os itens do teste de ciência da quarta série não podem ser tão fáceis que um aluno médio da quarta série responda todos os itens corretamente, ou não pode ser tão difícil que esse aluno responda todos os itens incorretamente. O índice de discriminação dos itens é usado para selecionar itens que diferenciem bem entre examinados com baixo e alto nível de proficiência. A TRI também fornece um indicador de precisão dos escores do teste: o erro padrão para cada examinado. Juntos, dificuldade e discriminação podem ser usadas para calcular o erro padrão da medida ou a confiabilidade do escore. Estes índices básicos fornecidos pela TRI tem analogias com os testes da teoria clássica (CTC).

Assim, a teoria de resposta ao item fornece uma estrutura útil para a resolução

de uma grande variedade de problemas de medição. Como a TRI é empregada através de um modelo matemático, a vantagem em utilizá-la depende fundamentalmente da escolha adequada do modelo através da adequação (ajuste) dos dados ao modelo escolhido e seus pressupostos.

1.1.2 Pressupostos da Teoria de Resposta ao Item

Segundo Reckase (2009), a metodologia da TRI define funções matemáticas que são usadas para relacionar a localização de uma pessoa em um espaço multidimensional com uma resposta correta para um item de um teste. A relação é medida pelas características do item (como a dificuldade, a discriminação e a probabilidade de chute), apresentadas por valores (parâmetros) que são estimados a partir dos dados de resposta ao item. O desenvolvimento da função matemática baseia-se em uma série de pressupostos.

O primeiro pressuposto é que as pessoas em um espaço de coordenadas cartesianas multidimensional não mudam durante o processo de respostas ao teste (uma questão respondida não pode auxiliar na resposta a outra questão).

O segundo pressuposto é de que as características de um item de um teste permanecem constantes sempre que o teste é aplicado. Por exemplo, um teste em inglês aplicado numa região que domina a língua não funciona tão bem em uma região onde a linguagem dominante é o espanhol, ou seja, há uma variação na população examinada.

Uma terceira suposição é que as respostas de uma pessoa para um item são independentes de suas respostas a outros itens do teste. Esta hipótese está relacionada com a primeira suposição. Não pode ser esperado que itens possam dar informações que podem melhorar o desempenho em itens posteriores. Da mesma forma, as respostas geradas por uma pessoa são assumidas não influenciar as respostas de outra pessoa.

Uma quarta suposição é de que a relação entre localizações no espaço multidimensional e as probabilidades de uma resposta correta para um item pode ser representada como uma função matemática contínua. Isto significa que, para cada local no espaço multidimensional, não há nenhuma descontinuidade. Este pressuposto

é importante para as formas matemáticas de modelos que representam as interações entre pessoas e itens de um teste.

Uma suposição final é que a probabilidade de resposta correta ao item aumenta, ou pelo menos não diminui, com o aumento da localização dos examinados nas coordenadas, em qualquer uma das dimensões. Isto é a suposição de ‘monotonicidade’ e parece razoável para os itens de teste projetados para a avaliação de habilidades e conhecimentos.

1.2 Modelos Unidimensionais da Teoria de Resposta ao Item

A ideia dos modelos unidimensionais é dispor de uma escala de medição que meça o traço latente. Uma escala arbitrária é definida e assume-se que, independente da habilidade do respondente, ela pode ser medida numa escala que tem um ponto central zero, uma unidade de medida igual a um, e um intervalo de valores possíveis variando de menos infinito a mais infinito. Assim, é possível verificar qual a habilidade de uma pessoa e comparar a de várias pessoas (e ordená-las). Apesar de a habilidade poder assumir valores de menos infinito a mais infinito, na prática costumam limitar de -3 a +3 (o valor encontrado na escala para cada respondente também é chamado de escore da pessoa). Reckase (2009) chama o escore da pessoa de continuum:

O conceito de continuum que está sendo utilizado aqui é similar ao conceito de “construção hipotética” utilizado na literatura da psicologia (MacCorquodale e Meehl 1948). Isto é, um continuum é uma escala ao longo da qual os indivíduos podem ser ordenados. Distâncias ao longo deste continuum têm sentido uma vez que uma origem para a escala e as unidades de medida são especificadas. Acredita-se que o continuum exista, mas não é diretamente observável. Sua existência é inferida a partir dos dados observados, neste caso, as respostas aos itens de um teste. O número de continuums necessários para descrever as diferenças entre as pessoas é assumido ser finito, mas grande. Em geral, o número de continuums em que um grupo de pessoas se diferem é muito grande e muito maior do que podia ser medido com qualquer teste atual. (Reckase, 2009).

Juntamente com a suposição de um parâmetro único de pessoa, a maioria dos modelos unidimensionais supõe que a probabilidade de selecionar ou produzir a resposta correta para um item aumenta com o aumento de θ .

Modelos matemáticos ajudam a fornecer um meio de quantificar os fenômenos de interesse. A teoria de Resposta ao Item propõe modelos matemáticos para características do indivíduo que não podem ser observadas diretamente (traço latente). As interações entre as características dos itens e das pessoas podem ser expressas de diversas formas, por infinitos modelos, dependendo das características dos dados. Alguns modelos serão citados a seguir:

1.2.1 Modelos Dicotômicos

O item é dito dicotômico ou não de acordo com a recodificação dada ‘as respostas. Quando o item é dicotômico, o indivíduo recebe um valor 1 para resposta correta ao item e para uma resposta incorreta recebe uma pontuação 0. Itens dicotômicos são também chamados de itens binários.

Os modelos dicotômicos são os mais utilizados. Para a maior parte dos itens com escores dicotômicos, é necessária uma função que relaciona os parâmetros da probabilidade de resposta correta de modo a que a hipótese de monotonicidade seja satisfeita. Isto é, com o aumento de θ , a forma funcional do modelo deve especificar que a probabilidade de resposta correta aumenta. Como só existem apenas duas categorias, os modelos dicotômicos mostram a probabilidade apenas do escore 1. A probabilidade do escore 0 é um menos a probabilidade do escore 1.

Modelo Logístico de um parâmetro

O modelo logístico de um parâmetro foi criado por Rasch (1960) e posteriormente descrito por um modelo logístico por Wright (1977). Além de ter a propriedade da monotonicidade, o modelo possui a vantagem da simplicidade. É definido por:

$$P(U_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-D(\theta_j - b_i)}} \quad (1.5)$$

Onde:

$i=1, \dots, I$ e $j=1, \dots, J$

b_i é o parâmetro de dificuldade do item i . Ele é medido na mesma escala da habilidade.

D é um fator de escala, constante e igual a 1. Utiliza-se o valor 1,7 quando deseja-se que a função logística forneça resultados semelhantes ao da função ogiva normal.

U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i , ou 0 quando o indivíduo j não responde corretamente ao item i .

Note que $P(U_{ij} = 1|\theta_j)$ pode ser vista como a proporção de respostas corretas ao item i dentre os indivíduos com habilidade θ_j . A relação existente entre $P(U_{ij} = 1|\theta_j)$ e os parâmetros do modelo pode ser visualizada pela figura 1.1 (a).

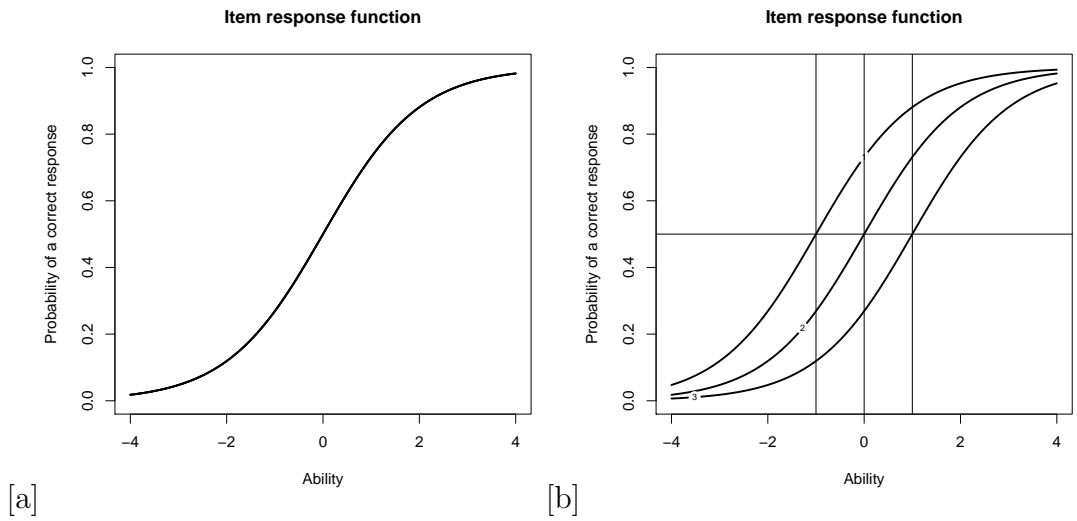


Figura 1.1: Curva Característica do Item: [a] item 1: $b=0$ [b] item 1: $b=-1$, item 2: $b=0$ e item 3: $b=1$

A suposição existente em um modelo logístico é que cada examinado que responde a um item do teste possui certa quantidade da habilidade subjacente. Assim, pode-se considerar que cada examinado tem um valor numérico (escore ou continuum) que coloca ele num lugar da escala de habilidade, simbolizado por θ . Em cada nível de habilidade, haverá certa probabilidade $P(\theta)$ de um examinado com habilidade θ dar uma resposta correta para o item. Quanto mais baixa a habilidade, mais baixa a probabilidade (mais $P(\theta)$ se aproxima de 0) e quanto mais alta a habilidade, mais alta será a probabilidade (mais $P(\theta)$ se aproxima de 1). Cada item em um teste terá a sua própria curva característica do item.

O modelo é denominado logístico de um parâmetro porque o único parâmetro considerado é a dificuldade do item (b). Apesar de receber a mesma denominação da Teoria Clássica, o parâmetro de dificuldade do item não é medido por uma proporção (valor entre 0 e 1), ele pode receber qualquer valor entre $-\infty$ a $+\infty$. Na teoria de resposta ao item, a dificuldade de um item descreve onde está a função do item ao longo da escala de habilidade. Para o modelo logístico de um parâmetro, todos os itens possuem o valor de b_i quando $P(\theta) = 0,5$. Observe na figura 1.1 (b) que todos os valores b produzem a mesma linha horizontal, com um valor de $P(\theta) = 0,5$.

Na figura 1.1 (b), a curva do lado esquerdo com $b = -1$ representa um item mais fácil, porque a probabilidade de resposta correta é alta para os examinandos de baixa habilidade e se aproxima de 1 para examinados com alta habilidade. A curva central representa um item de dificuldade média com $b = 0$, porque a probabilidade de resposta correta é baixa nos níveis mais baixos de habilidade, em torno de 0,5 no meio da escala de habilidade e próximo de 1 nos mais altos. Observe que quando a dificuldade do item é inferior a 0, a maior parte da curva característica item tem uma probabilidade de resposta correta, que é maior do que 0,5. Quando a dificuldade do item é maior que 0, a maior parte da curva característica do item tem uma probabilidade de resposta correta inferior a 0,5. O parâmetro b divide a função ao meio e está centrado entre a assíntota inferior e a assíntota superior da função.

Modelo Logístico de dois parâmetros

O modelo logístico de dois parâmetros é definido por:

$$P(U_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (1.6)$$

Onde:

$i=1, \dots, I$ e $j=1, \dots, J$

U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i , ou 0 quando o indivíduo j não responde corretamente ao item i .

a_i é o parâmetro de discriminação (ou de inclinação) do item i , com valor proporcional ‘a inclinação da curva características do Item - CCI no ponto b_i ’.

b_i é o parâmetro de dificuldade do item i . Ele é medido na mesma escala da habilidade.

D é um fator de escala, constante e igual a 1. Utiliza-se o valor 1,7 quando deseja-se que a função logística forneça resultados semelhantes ao da função ogiva normal.

O modelo logístico de dois parâmetros se difere do de um parâmetro por ter o parâmetro de discriminação (a) incluso na fórmula. O parâmetro a é proporcional à derivada da tangente da curva no ponto de inflexão. Este parâmetro não é uma correlação (não recebe um valor entre -1 e 1 como na teoria clássica dos testes). Ele pode receber qualquer valor entre $+\infty$ a $-\infty$. Porém, itens com valores de a negativos não são esperados nesses modelos, pois a curva mudaria de crescente para decrescente e indicaria que a probabilidade de responder corretamente ao item diminui com o aumento da habilidade. Quanto mais baixo o valor de a menor o poder que o item tem de diferenciar os indivíduos com baixa habilidade dos indivíduos com alta habilidade. Valores muito altos de a indicam itens com curvas muito “ingremes”, que discriminam os alunos em basicamente dois grupos: os que possuem habilidades abaixo do valor do parâmetro b e os que possuem habilidades acima do parâmetro b .

Enquanto a maioria dos itens do teste deve discriminar de forma positiva, ou seja, a probabilidade de resposta correta aumenta com o aumento do nível de habilidade, alguns itens têm discriminação negativa. Nesses itens, a probabilidade de resposta correta diminui a medida que o nível de habilidade aumenta. Isto pode ocorrer por erro na codificação dos itens (respostas incorretas ganharem valores positivos), ou há alguma desinformação predominante entre os indivíduos com alta habilidade. De qualquer forma, esses itens necessitam de alguma atenção.

A figura 1.2 (a) ilustra três itens com parâmetro de dificuldade médio $b = 0$ e diferentes parâmetros de discriminação: 0.2, 1 e 3, para os itens 1, 2 e 3, respectivamente. Observe que o item 1 (com $a = 0.2$) possui baixo valor de discriminação e a diferença é pequena entre a probabilidade de acerto $P(U_{ij} = 1|\theta_j)$ dos diferentes níveis de habilidades (θ). Já no item, com $a = 3$, os indivíduos com habilidades um pouco abaixo de $b = 0$ possuem probabilidade bem menor do que aqueles que estão localizados acima do ponto $b = 0$, ou seja, este item discrimina bastante entre os respondentes com proficiência abaixo do índice de dificuldade do item daqueles que

estão com proficiência acima do valor de b .

Na figura 1.2 (b), o índice de dificuldade dos itens 2 e 4 é mais baixo, $b = -1$, do que os itens 1 e 3 ($b = 1$). Os itens 2 e 4 (assim como 1 e 3) se diferenciam quanto ao índice de discriminação do item, onde os itens 2 e 3 são mais discriminativos, pois possuem maior diferença entre a probabilidade de acerto $P(U_{ij} = 1|\theta_j)$ dos indivíduos que estão com habilidades θ abaixo e acima de b .

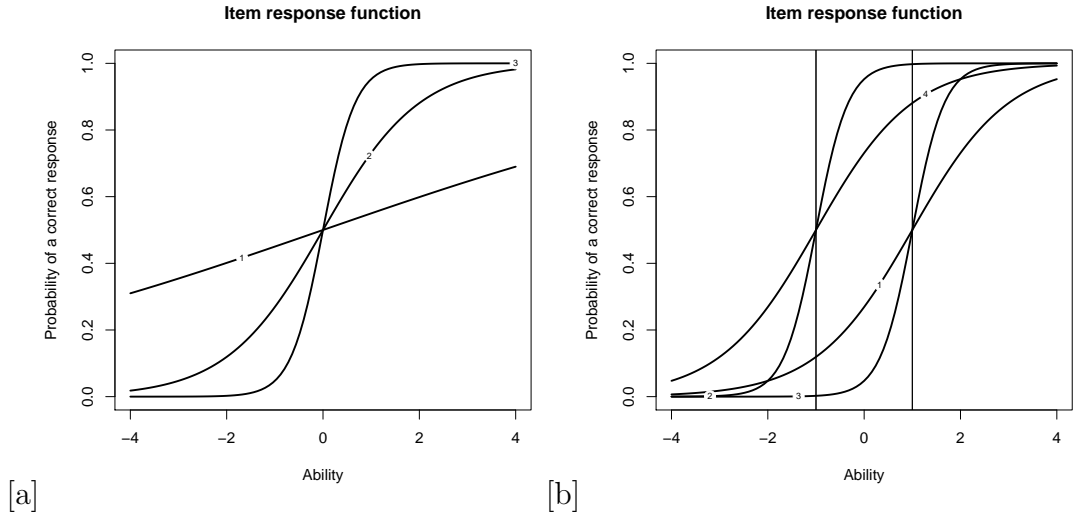


Figura 1.2: Curva Característica do Item : [a] item 1: $a = 0$; 2 e $b = 0$, item 2: $a = 1$ e $b = 0$, item 3: $a = 3$ e $b = 0$ [b] item 1: $a = 1$ e $b = 1$, item 2: $a = 3$ e $b = -1$, item 3: $a = 3$ e $b = 1$, item 4: $a = 1$ e $b = -1$

Modelo Logístico de três parâmetros

O modelo logístico de três parâmetros é definido por:

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (1.7)$$

Onde:

$$i=1, \dots, I \text{ e } j=1, \dots, J$$

U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i , ou 0 quando o indivíduo j não responde corretamente ao item i .

a_i é o parâmetro de discriminação (ou de inclinação) do item i , com valor proporcional ‘a inclinação da curva características do Item - CCI no ponto b_i .

b_i é o parâmetro de dificuldade do item i . Ele é medido na mesma escala da habilidade.

c_i é a probabilidade do indivíduo i acertar o item ao acaso (chute).

D é um fator de escala, constante e igual a 1. Utiliza-se o valor 1,7 quando deseja-se que a função logística forneça resultados semelhantes ao da função ogiva normal.

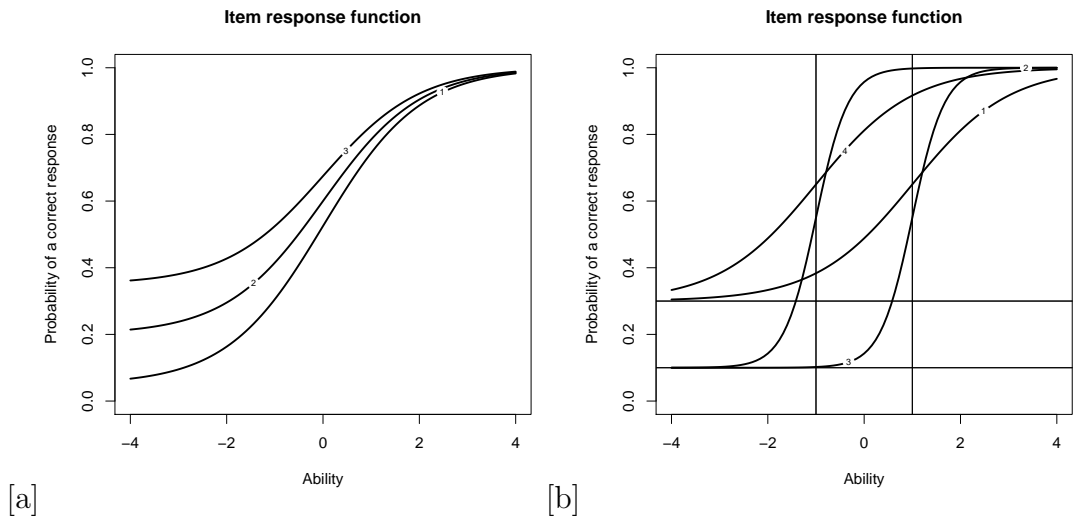


Figura 1.3: Curva Característica do Item: [a] item 1: $a = 1$ e $b = 0$, item 2: $a = 1$ e $b = 0$, item 3: $a = 3$ e $b = 0$ [b] item 1: $a = 1$, $b = 1$ e $c = 0:3$; item 2: $a = 3$, $b = -1$ e $c = 0:1$; item 3: $a = 3$ e $b = 1$ e $c = 0:1$; item 4: $a = 1$, $b = -1$ e $c = 0:3$

É importante notar que, por definição, o valor de c não varia em função do nível de habilidade. Assim, os examinandos com habilidade menor e maior têm a mesma probabilidade de obter o item correto chutando. O parâmetro c tem um alcance teórico de 0 até 1 (já que é uma probabilidade), mas, na prática, os valores acima de 0,35 não são considerados aceitáveis. Um efeito da utilização do parâmetro do chute recai sobre a definição do parâmetro de dificuldade. Nos modelos logísticos de um e dois parâmetros, o valor de b corresponde ao valor na escala de habilidade (θ) no qual a probabilidade de resposta correta é 0,5. No modelo de três parâmetros, a definição do parâmetro de dificuldade é alterada. O valor fixado para b está situado na

escala de habilidade (θ) onde a probabilidade de resposta correta está na metade do valor entre c e 1, isto é, $(1 + c)/2$. Isto porque o limite inferior da curva característica item é o valor de c , em vez de zero.

O parâmetro de discriminação, nesse modelo, ainda pode ser interpretado como sendo proporcional a inclinação da curva característica do item no ponto de $\theta = b$. No entanto, sob o modelo de três parâmetros, a inclinação da curva característica do item em $\theta = b$ é dada por $(1 - c) / 4$. Observe que, com a inclusão de c , a função deixa de ser logística, apesar de ser chamada assim por convenção.

1.2.2 Modelos Não Dicotômicos

A maioria dos testes utilizados em teoria de resposta ao item é composta de itens de múltipla escolha. Enquanto os modelos dicotômicos consideram as respostas de uma questão de múltipla escolha em certo e errado, codificando em 1 para a escolha correta e zero para qualquer outra escolha do item, os modelos não dicotômicos levam em conta toda a informação contida na questão, tentando obter mais informação das respostas dos indivíduos do que simplesmente se eles deram respostas corretas ou incorretas. Itens politômicos possuem modelos específicos, onde não consideram apenas se o indivíduo respondeu corretamente ou não, mas também levam em conta qual foi a resposta dada por ele. Modelos nominais podem ser aplicados a todas as categorias de resposta escolhidas em um teste de múltipla escolha e não considera a ordenação das alternativas. Modelos Graduais levam em conta todas as categorias de resposta e também a ordenação das alternativas (como numa escala likert, por exemplo).

Modelo de Resposta Nominal (Nominal Categories Model)

O Modelo de Resposta nominal foi baseado no modelo logístico de dois parâmetros e pode ser aplicado a todas as categorias de resposta escolhidas em um teste de múltipla escolha. O propósito desse modelo é melhorar a precisão da habilidade estimada utilizando toda a informação contida nas respostas dos indivíduos, e não apenas se o item foi respondido corretamente ou não. O modelo assume que a probabilidade com que um indivíduo j selecionar uma particular opção k (de m_i opções existentes)

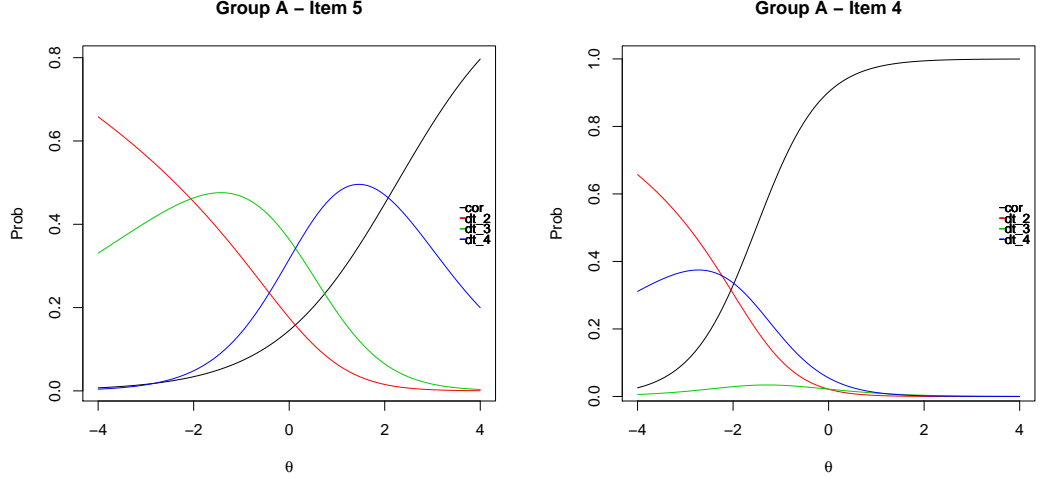


Figura 1.4: Curva Característica de dois Itens para o Modelo Nominal

do item i é dada por:

$$P_{i,k}(\theta_j) = \frac{e^{a_{j,k}^+(\theta_j - b_{i,j}^+)}}{\sum_{h=1}^{m_i} e^{a_{i,k}^+(\theta_j - b_{i,k}^+)}} \quad (1.8)$$

com $i = 1, 2, \dots, I$, $j = 1, 2, \dots, n$ e $k = 1, 2, \dots, m_i$. Em cada θ_j , a soma das probabilidades sobre as m_i opções é igual a 1. A quantidade $a_{i,k}^+$ é um parâmetro do item i relacionados a k -ésima opção. O modelo assume que não há nenhuma ordenação a priori das opções de resposta.

Modelo de Resposta Gradual (Graded Response Model)

Diferentemente do modelo nominal, o modelo de resposta gradual de Samejima (1969) assume que existe uma ordenação entre as categorias de respostas de um item. Se os escores das categorias de um item i são ordenados do menor para o maior (como em uma escala Likert, por exemplo) não é adequado codificá-las em certo e errado. Considere a categoria k , $k=0,1, \dots, m_i$ onde $(m_i + 1)$ é o número de categorias do i -ésimo item. A probabilidade de um indivíduo j escolher uma particular categoria ou outra mais alta do item i pode ser dada por uma extensão do modelo logístico de 2 parâmetros:

$$P_{i,k}^+ = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}} \quad (1.9)$$

$b_{i,k}$ é o parâmetro de dificuldade da k -ésima categoria do item i . Os demais parâmetros já foram definidos nos outros modelos. Diferentemente dos modelos dicotômicos, a discriminação de uma categoria específica de resposta depende tanto do parâmetro de inclinação (como nos modelos dicotômicos), comum a todas as categorias do item, quanto da distância das categorias de dificuldade adjacentes. É importante se assegurar que a ordenação entre os níveis de dificuldade das categorias de um dado item exista, de acordo com a classificação dos seus escores, ou seja, $b_{i,1} \leq b_{i,2} \leq \dots \leq b_{i,m_i}$.

A probabilidade de um indivíduo j receber um escore k no item i é dada por:

$$P_{i,k}(\theta_j) = P_{i,k}^+(\theta_j) - P_{i,k+1}^+(\theta_j)$$

se $P_{i,0}^+(\theta_j)$ e $P_{i,m_i+1}^+(\theta_j)$ forem definidos de forma que:

$$P_{i,0}^+(\theta_j) = 1 \text{ e } P_{i,m_i+1}^+(\theta_j) = 0$$

Então $P_{i,0}(\theta_j) = P_{i,0}^+(\theta_j) - P_{i,1}^+(\theta_j) = 1 - P_{i,1}^+(\theta_j)$ e

$$P_{i,m}(\theta_j) = P_{i,m}^+(\theta_j) - P_{i,m+1}^+(\theta_j) = P_{i,m}^+(\theta_j)$$

Então temos que:

$$P_{i,k}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k+1})}} \quad (1.10)$$

Note que em um item com $(m_i + 1)$ categorias, m_i valores de dificuldade necessitam ser estimados, além do parâmetro de inclinação do item. Assim, para cada item, o número de parâmetros a serem estimados será dado pelo seu número de categoria de resposta. Se, por exemplo, tivermos um teste com I itens, cada um com $(m_i + 1)$ categorias de resposta, teremos então $\left[\sum_{i=1}^I m_i + I \right]$ parâmetros a serem estimados.

Modelo de Escala Gradual (Rating Scale Model)

É um caso particular do modelo de resposta gradual de Samejima. Esse modelo também é adequado para itens com categorias de resposta ordenadas. No entanto, é feita a suposição de que os escores das categorias são igualmente espaçados.

Este modelo, proposto por Andrich (1978), é dado por:

$$P_{i,k}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i + d_k)}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_i + d_{k+1})}} \quad (1.11)$$

com $i = 1, 2, \dots, I$, $j = 1, 2, \dots, n$ e $k = 0, 1, \dots, m$.

onde: b_i é agora o parâmetro de locação do item i e d_k é o parâmetro de categoria.

Como $P_{i,j}^+(\theta_j) - P_{i,k+1}^+(\theta_j) \geq 0$.

Devemos ter: $d_1 \geq d_2 \geq \dots \geq d_m$

A maior difetença entre o Modelo de Resposta Gradual e o Modelo de Escala Gradual está na hipótese de nesse último os escores das categorias de resposta devem ser equidistantes. Assim, no Modelo de Escala Gradual o parâmetro $b_{i,k}$ é decomposto em um parâmetro b_i de locação do item e num parâmetro de categoria d_k , isto é: $b_{i,k} = b_i - d_k$.

Cabe ressaltar que os parâmetros de categoria d_k não dependem do item, isto é, são comuns a todos os itens do teste. Logo, se os itens que compõem o teste tiverem suas próprias categorias de resposta, que podem diferir no número, então este modelo não é adequado. Em um teste composto por itens com $(m + 1)$ categorias de resposta cada um, m parâmetros de categoria necessitam ser estimados, além dos parâmetros de inclinação e de locação de cada item. Logo, se o teste tiver I itens, teremos $[2I+m]$ parâmetros a serem estimados.

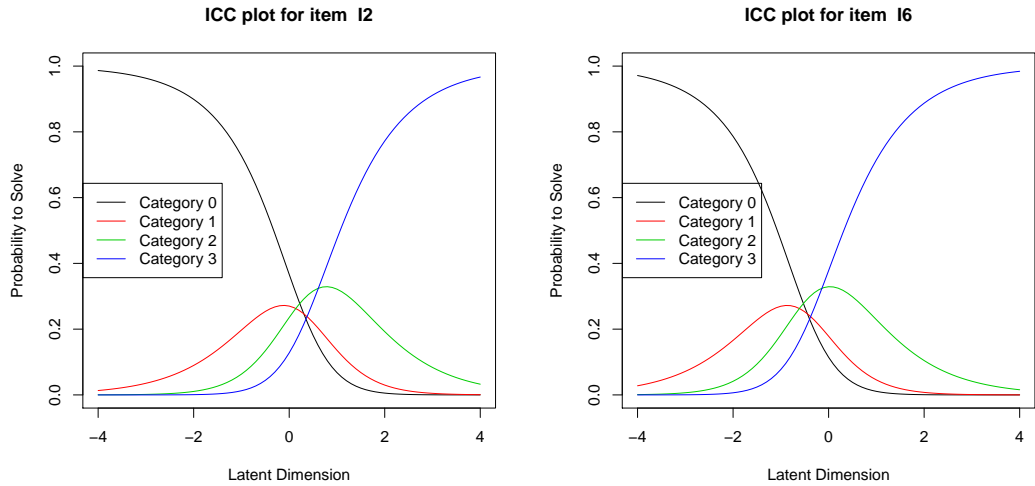


Figura 1.5: Curva Característica do Item para o Modelo de Escala Gradual

Modelos envolvendo mais de uma População

Alguns modelos foram desenvolvidos para serem aplicados quando o teste envolve mais de uma população. É pelo processo de amostragem do estudo que se identifica quantas (e quais) são as populações envolvidas. Por exemplo, se o estudo em questão é dos alunos de uma determinada série de ensino de uma determinada UF, toma-se uma amostra de alunos dessa população (e será composta por alunos do diurno e do noturno). Nesse caso, tem-se apenas uma população de respondentes. Se o estudo considera os alunos do diurno diferentes dos alunos do noturnos, seriam então necessárias duas amostras e seriam então duas populações a serem avaliadas. Exemplos que usualmente são consideradas como populações distintas: séries distintas; turnos distintos; mesma série em anos distintos, etc. Assim, existem modelos de teoria de resposta ao item diferentes quanto ao número de populações estudadas.

Geralmente, indivíduos de diferentes populações não são submetidos aos mesmos itens a serem testados. No entanto, para comparar diferentes populações é necessário que haja pelo menos alguns itens comuns entre elas. Assim, sendo I o número total de itens distintos apresentados, o modelo proposto por Bock e Zinmowski (1997) é dado por:

$$P(U_{ijk} = 1|\theta_{jk}) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_{jk} - b_i)}} \quad (1.12)$$

Onde:

U_{ijk} é uma variável dicotômica que assume os valores 1, quando o indivíduo j da população k responde corretamente o item i , ou 0 quando o indivíduo j não responde corretamente ao item i .

θ_{jk} representa a habilidade do j -ésimo indivíduo da população k .

$P(U_{ijk} = 1|\theta_{jk})$ é a probabilidade de um indivíduo j da população k , com habilidade θ_{jk} , responder corretamente ao item i . Os demais parâmetros são descritos como os anteriores.

Estimação dos Parâmetros da Teoria de Resposta ao Item

Na Teoria de Resposta ao Item unidimensional, a probabilidade de uma resposta correta a um determinado item depende somente da habilidade do indivíduo e dos

parâmetros que caracterizam o item. Mas, apenas as respostas ao item são conhecidas. Assim, tem-se um problema de estimação que envolve dois tipos de parâmetros: os parâmetros dos itens e o parâmetro da habilidade do indivíduo. Na TRI, o processo de estimação dos itens se chama calibração.

A TRI nos possibilita analisar individualmente os itens de um teste fornecendo parâmetros referentes aos itens e aos indivíduos. Bem interpretados, esses parâmetros fornecem uma grande quantidade de informação necessária para a análise dos itens, identificando itens por sua dificuldade e por seu poder de discriminação entre os indivíduos de maior ou menor nível de habilidade.

Existem três situações para estimativas dos parâmetros: conhecer as habilidades e estimar apenas os parâmetros do item - tem apenas caráter teórico e é solucionado usando o método da máxima verossimilhança; conhecer os parâmetros dos itens e estimar apenas as habilidades - Nesse caso a solução é dada empregando o método da máxima verossimilhança ou métodos bayesianos, ambos através da aplicação de procedimentos iterativos, como, por exemplo, o método de Newton-Raphson ou Scoring de Fisher; e não conhecer nem os parâmetros dos itens e nem as habilidades. Quando ambos os parâmetros devem ser estimados também pode ser usado o método da máxima verossimilhança há duas abordagens usuais: estimação conjunta (dos parâmetros dos itens e das habilidades), ou em duas etapas, primeiro a estimação dos parâmetros dos itens e, posteriormente, das habilidades.

O ajuste do modelo escolhido aos dados deve ser comprovado pelos parâmetros estimados e começa na construção do instrumento de medida. Trata-se da definição rigorosa da dimensão que se pretende avaliar, seguida da elaboração dos itens que representem adequadamente essa dimensão.

O segundo passo é a aplicação dos itens a uma amostra representativa da população para a qual se constrói o instrumento, com a finalidade de verificar a unidimensionalidade dos itens. Nessa fase, escolhe-se o modelo que melhor ajusta-se aos dados, por exemplo, se a discriminação dos itens não é constante, escolhe-se o modelo de dois parâmetros. Se houver probabilidade de acerto ao acaso, os dados se ajustam melhor ao modelo de tres parâmetros. Por exemplo, se os itens são de múltipla escolha, é possível o acerto ao acaso, entretanto, se utilizar um modelo de um parâmetro, o resíduo gerado pode ser tão inexpressivo que o uso desse modelo

poderá ser apropriado nesse caso.

A partir das respostas dos sujeitos aos itens, se inicia a estimação dos parâmetros. Um dos métodos usados é chamado Máxima Verossimilhança no qual os valores são aqueles que maximizam a probabilidade de ocorrência dos dados. Os valores estimados são aqueles que se fazem mais plausíveis para os dados obtidos. O problema da estimação dos parâmetros dos itens na TRI é que conhece-se apenas as respostas dadas aos itens. Como os parâmetros são desconhecidos, é necessário realizar um processo iterativo, como, por exemplo, assumir valores iniciais hipotéticos para os parâmetros dos itens, estimar as habilidades dos sujeitos, considerar esses novos valores provisórios para reestimar os parâmetros dos itens de maneira um pouco mais acurada e assim sucessivamente até que não se consiga melhorar mais os ajustes das curvas teóricas aos dados empíricos.

Os procedimentos de estimação para esse método podem assumir duas formas distintas. A primeira consiste em fixar valores para os parâmetros dos itens e se realizar sucessivas estimações para diferentes valores possíveis de θ , até que o valor que melhor explique a ocorrência do padrão empírico de respostas seja encontrado. São testadas todas as possibilidades. Esta forma é conhecida por estimação condicional como uma referência à condição do conhecimento prévio dos parâmetros dos itens. A segunda forma consiste em se calcular ao mesmo tempo tanto os parâmetros dos itens quanto o θ de cada sujeito e é conhecida como estimação conjunta. Primeiramente, se estima os valores de θ para cada sujeito, supondo os parâmetros dos itens tal como na estimação condicional, indicando um valor inicial, após os valores de θ conhecidos se retorna calculando os parâmetros dos itens para aqueles valores. A diferença entre os procedimentos é que, no primeiro caso, se estimam as habilidades de todos os sujeitos para depois se corrigir o valor dos parâmetros e no segundo caso, as estimações da habilidade e dos parâmetros dos itens são feitas concomitantemente (Muñiz, 1990).

Alguns passos para ilustrar a estimação de parâmetros utilizando o método de Máxima Verossimilhança: sejam n indivíduos que respondem I itens de um teste. Supoem-se que a única informação disponível seja as respostas aos itens. Então o primeiro passo consiste em separar os sujeitos em grupos ao longo de uma escala de habilidade hipotética, cada grupo tem Z sujeitos de habilidades iguais. A probabilidade de os sujeitos de cada grupo responderem adequadamente a um item específico

será dada pelo quociente entre o número de sujeitos que realmente acertaram ao item e o número total de sujeitos daquele grupo. Dessa forma as probabilidades de acerto em cada nível de habilidade ao longo da escala podem ser calculadas, isto é, tem-se uma curva empírica para cada item. A partir disso tenta-se manipular os parâmetros do item, produzindo uma curva teórica que mais se aproxime da empírica. O processo de estimação dos parâmetros se encerra quando os valores estimados convergirem, ou seja, quando a partir de n interações não se consegue produzir mais melhorias na reprodução dos dados empíricos por meio das variações nos valores dos parâmetros dos itens (Wright e Stone, 2004; Baker, 2001; Muñiz, 1990).

Após a estimação dos parâmetros do modelo os valores encontrados devem ser confrontados com os resultados empíricos, ou seja, as respostas dos sujeitos, e verificar se existem diferenças estatisticamente significativas. Existem vários procedimentos estatísticos usados para a comprovação do ajuste do modelo aos dados, mas nenhum é totalmente satisfatório, o que acaba por se tornar o ponto fraco da TRI no seu momento atual de desenvolvimento (Muñiz, 1990; Baker, 2001; Wright e Stone, 2004).

Informações mais detalhadas sobre a estimação dos parâmetros podem ser encontradas em [Baker (1987), Baker (1992), Lord (1980), Birnbaum (1968), Swaminathan e Gifford (1983), Hambleton e Swaminathan (1985), Hambleton et al. (1991), Hambleton e Van Der Linden (1996)].

1.2.3 Métodos de Estimação dos Parâmetros

Os principais pressupostos para estimação de parâmetros da Teoria de Resposta ao Item são: Independência entre as respostas dos indivíduos e independência entre as respostas dadas aos itens, para uma proficiência específica (local ou condicional). Nessa seção serão descritos alguns métodos de estimação para uma população, dado que os pressupostos são seguidos:

Estimação por Maxima Verossimilhança Conjunta

$$L(\theta, \xi) = P(U|\theta, \xi) = \prod_{j=1}^N \prod_{i=1}^I P_{ij}^{u_{ij}} (1 - P_{ij})^{1-u_{ij}} \quad (1.13)$$

$$\text{Log}L = \sum_{j=1}^N \sum_{i=1}^I u_{ij} \log P_{ij} + (1 - u_{ij}) \log(1 - P_{ij}) \quad (1.14)$$

Onde $U = (u_{ij})$ é a matriz das respostas ($N \times I$) e ξ é o vetor ($q \times 1$) dos parâmetros dos itens. Para o modelo logístico de 3 parâmetros, $q=3$

É necessário encontrar os valores de θ e ξ que maximizam $\log L$ através de uma técnica iterativa, por exemplo, Newton-Raphson. São necessárias as derivadas parciais de primeira e segunda ordem de $\log L$ com respeito a θ e ξ . Indeterminação: existem diferentes valores de θ e b que fornecem o mesmo valor de P_{ij} . Uma solução: θ s com média 0 e desvio padrão 1, escala (0,1). Uma outra técnica é a de verossimilhança marginal.

Estimação por Máxima verossimilhança marginal

A idéia básica é “libertar o processo de estimação dos parâmetros dos itens de sua dependência de θ ”. O primeiro passo é a estimação dos parâmetros dos itens. O segundo passo é: assumindo que as estimativas dos parâmetros dos itens são seus verdadeiros valores, estimados os θ s.

$$P(U_j | \theta, \xi) = \prod_{i=1}^{u_{ij}} P_{ij}^{u_{ij}} (1 - P_{ij})^{(1 - u_{ij})} \quad (1.15)$$

$$P(U_j | \theta, \eta) = \oint_{\mathbb{R}} P(u_j | \theta, \xi) g(\theta, \eta) d\theta \quad (1.16)$$

$$L(\xi, \eta) = P(U | \xi, \eta) = \prod_{j=1}^n P(U_j | \xi, \eta) \quad (1.17)$$

Onde: $g(\theta | \eta)$ é a distribuição de θ , com parâmetros $\eta = (\mu, \sigma^2)^1$. Em geral, consideramos a normal padrão ($\mu = 0$ e $\sigma = 1$).

As estimativas dos parâmetros dos itens são os valores de ξ que maximizam $L(\xi, \eta)$. O Algoritmo EM pode ser utilizado, onde: U e θ são os dados completos, e u é dado observado. Assumindo ξ “conhecido”, retorna-se para $L(\xi, \theta) = L(\theta)$, maximizando para θ .

Estimação Bayesiana

Diferente da estimação por máxima verossimilhança, a estimação bayesiana fornece estimativas para todos os itens com $u = 1$ ou $u = 0$. Geralmente, as funções a ‘priori’ utilizadas são:

- Distribution a priori para a : Lognormal
- Distribuição a priori para b : Normal
- Distribuição a priori c : Beta

Duas ou mais populações

Caso 1: Estimação para cada população em separado: requer uma “equalização a posteriori” para ter todos os resultados na mesma escala (métrica).

Caso 2: Estimação envolvendo todas as populações ao mesmo tempo. Enfoque de Grupos Múltiplos: Estabelecemos uma das populações (grupos) como a referência, e obtemos todos os resultados na mesma escala. Por exemplo, estabelecemos a escala (0,1) para a população 1, e todos os resultados das outras populações estarão na mesma escala.

1.2.4 Teoria de Resposta ao Item Multidimensional

Na TRI, a habilidade é observada através de um conjunto de itens, onde cada item deve medir algum aspecto do construto de interesse. Essa relação é de tal forma que quanto maior a habilidade, maior a probabilidade de acerto no item. Porém, mais de uma habilidade podem interagir com os itens respondidos. Quando apenas uma habilidade é considerada no modelo, esse denomina-se unidimensional e quando há mais de uma habilidade o modelo chama-se multidimensional.

Os modelos atuais de TRI consideram que as avaliações medem apenas um traço latente (habilidade), ou seja, podem ser representado em apenas uma dimensão. Mas, pela própria construção dos itens, isto pode não ser verdade (como exemplo, a redação do ENEM ¹ que é estruturada em 5 competências). Por isso, técnicas

¹No caso da redação, cinco competências são avaliadas: I- Demonstrar domínio da norma padrão da língua escrita. II- Compreender a proposta de redação e aplicar conceitos das várias áreas de

multidimensionais são propostas, através de análises fatoriais, a fim de verificar a dimensionalidade dos testes aplicados.

A utilização de Modelos Multidimensionais da TRI iniciou-se a partir dos anos 70, baseada nos trabalhos de Horst (1965), McDonald (1967), Muthén (1978), Samejima (1974), entre outros. A história e o desenvolvimento destes modelos baseiam-se principalmente em duas áreas: a análise factorial (Horst, 1965; Christoffersson, 1975; Muthén, 1978; McDonald, 1967 e 1985; Bock e Aitkin, 1981 e Bejar, 1977) e a extensão dos modelos de resposta ao item unidimensionais (Rasch, 1962; Lord e Novick, 1968 e Samejima, 1974).

Dimensionalidade e independência local

O valor dado à proficiência (continuum) de um determinado indivíduo, que pode ser definido através de um conjunto de respostas ao item, depende da variação da amostra de indivíduos examinados e das características do item do teste. Os itens devem ser construídos de modo que a determinação da localização de pessoas em um espaço multidimensional consiga diferenciar essas pessoas.

Supõem-se que: ou a pessoa que o acerta ao item tem todas as habilidade e conhecimentos necessários para respondê-lo corretamente ou a pessoa é deficiente em algum componente do item. Esse componente poderia ser a habilidade de leitura ou conhecimento do vocabulário, ou o conhecimento do processo de teste de múltipla escolha com itens ou até mesmo o tempo que o aluno tem para realizar o teste, pois se o tempo foi curto e a resposta pode nem mesmo ser respondida. Para a construção do item, espera-se que esse componente tenha a ver com o construto a ser medido. Por exemplo, se está medindo a proficiência em matemática espera-se que o indivíduo erre ou acerte de acordo com seus conhecimentos em matemática. A princípio, o respondente precisa entender a linguagem do teste, o formato da pergunta, o significado das palavras que são alternativas de resposta, entre outras coisas. Por isso, os itens devem ser cuidadosamente desenvolvidos para não serem sensíveis a

conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo- argumentativo III- Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista. IV- Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da argumentação. V- Elaborar proposta de solução para o problema abordado, respeitando os valores humanos e considerando a diversidade sociocultural.

diferenças irrelevantes.

Os modelos mais utilizados são os unidimensionais, por não serem tão complexos como os modelos multidimensionais, e assumem que há apenas uma habilidade responsável pelas respostas ao item. Parece pacífico assumir que qualquer desempenho humano é sempre multideterminado ou multimotivado, dado que mais de um construto entre na execução de qualquer tarefa. Então, para admitir a unidimensionalidade, aceita-se que há um fator dominante responsável pelo conjunto de itens de um teste. E esse fator é o que se supõe estar sendo medido pelo teste. Porém, na própria construção dos itens, ou pela finalidade da aplicação, não existe apenas uma habilidade a ser testada, ou seja, não devem ser considerados unidimensionais (como o caso do Exame Nacional do Ensino Médio (ENEM), que foi estruturado segundo uma matriz de 5 competências).

Além das características de dificuldade e discriminação do item apresentadas tanto na TCT e na TRI (e de probabilidade de acertar ao acaso, em alguns modelos da TRI), o pressuposto de que o conjunto dos itens representam apenas o construto a ser avaliado deve ser questionada. Esse pressuposto chama-se unidimensionalidade, isto é, a homogeneidade do conjunto de itens que supostamente devem estar medindo um único traço latente. Isso porque ao descrever as interações entre as pessoas e os itens de um teste, deve-se considerar que as pessoas variam de acordo com as suas características. Então, se por um lado algumas dessas características (ou habilidades) são importantes para testar o desempenho em um item específico do teste, por outro lado os itens do teste podem exigir múltiplas habilidades para chegar a resposta correta (Reckase, 2009).

Se o objetivo é o de determinar a habilidade do indivíduo em um teste várias conhecimentos e fatores devem ser considerados, como por exemplo: o conhecimento da linguagem utilizada pra resolver o teste (se o objetivo é testar a habilidade em matemática, apenas, um teste com questões de interpretações pode conter um vocabulário que o respondente ignora e, apesar de ter conhecimento matemático para acertar a questão ele pode responder incorretamente); a tendência de adivinhar ('chutar') quando se está em dúvida sobre as respostas; a rapidez em responder as questões (se um teste for muito longo e o prazo para respondê-lo for curto, pode acontecer de alguns respondentes com conhecimento necessário não consiga demonstrar o seu

conhecimento, deixando questões sem respostas); o estado emocional do respondente pode ser uma dimensão a mais ao testar seu conhecimento (muitos respondentes podem ter conhecimento das questões do teste e por nervosismo não conseguem resolvê-las); a homogeneidade da amostra (se um teste de matemática é aplicado aos alunos do ENEM², aqueles que ainda não terminaram o ensino médio se diferem daqueles que concluíram e mais ainda dos que já cursaram o ensino superior, assim, uma questão de cálculo, por exemplo, será mais vantajosa para aqueles que já fizeram um curso de ciências exatas); dentro de um mesmo conhecimento testado podem existir diversas habilidades (num teste de matemática, por exemplo, pode está sendo testado Cálculo, Geometria Analítica, Vetores, Desenho, Cálculo Numérico, Álgebra, Física entre outras); dentro de um mesmo item pode estar sendo medido várias habilidades (reconhecer, no contexto social, diferentes significados e representações dos números e operações naturais, inteiros, racionais ou reais; Identificar padrões numéricos ou princípios de contagem; Resolver situação-problema envolvendo conhecimentos numéricos, etc):

“Um teste consiste em coleções de itens. Cada um dos itens do teste é complexo por sua própria natureza. As pessoas que fazem um teste igualmente consistem nos mais diversos indivíduos. Mesmo gêmeos idênticos irão mostrar algumas diferenças em seus conhecimentos e habilidades, porque suas experiências de vida não são exatamente o mesmo depois de seu nascimento. As interações entre examinandos e os itens de um teste em resultam em um conjunto de respostas que representam processos muito complexos.(Reckase, 2009)”.

Além da dimensionalidade, outro pressuposto a ser verificado é o de independência local, ou independência condicional, onde assume-se que para uma dada habilidade as respostas aos diferentes itens da prova são independentes. Essa suposição é fundamental para estimação dos parâmetros do modelo. Como unidimen-

²Segundo o ministério da educação, Podem participar do Enem 2013, todos aqueles que concluíram ou não o Ensino Médio e que possuam CPF e documento de identidade. A finalidade primordial do Enem é a avaliação do desempenho escolar e acadêmico ao fim do Ensino Médio. As informações obtidas a partir dos resultados do Enem são utilizadas para acompanhamento da qualidade do Ensino Médio no País, na implementação de políticas públicas, criação de referência nacional para o aperfeiçoamento dos currículos do Ensino Médio, desenvolvimento de estudos e indicadores sobre a educação brasileira e estabelecimento de critérios de acesso do participante a programas governamentais. O ENEM serve, também, para a constituição de parâmetros para a auto-avaliação do participante, com vistas à continuidade de sua formação e à sua inserção no mundo de trabalho.

sionalidade implica independência, verificar a unidimensionalidade já é suficiente. Hambleton, Swaminathan e Roger discutem as consequências da utilização de modelos inadequados aos dados e apresenta métodos para verificação do ajuste e dos pressupostos do modelo utilizados. Um bom ajuste pode garantir a obtenção de itens e habilidades invariantes.

“Independência local significa que, quando as habilidades que influenciam o desempenho no teste são mantidas constantes, as respostas dos examinados para dois itens quaisquer são estatisticamente independentes. Em outras palavras, após considerar as habilidades dos examinados nenhuma relação existe entre as respostas dos examinados para diferentes itens. Simplificando, isso significa que as habilidades especificadas no modelo são os únicos fatores que influenciam as respostas dos examinados aos itens do teste. Este conjunto de habilidades representa o espaço latente completo. Quando a suposição de unidimensionalidade detém, o espaço latente completo é composto de apenas uma habilidade. (Hambleton, Swaminathan e Roger, 1991)”.

1.2.5 Modelos Multidimensionais da Teoria de Resposta ao Item

A TRI multidimensional pode ser considerada um tipo de análise fatorial para dados observados não contínuos (dicotômicos ou politômicos). Porém, na análise fatorial procura-se um pequeno número de fatores capazes de explicar a variabilidade das variáveis observadas, ou seja, busca-se a redução da dimensão. Enquanto que na TRI busca-se determinar as características dos indivíduos e dos itens que influenciam nas respostas observadas.

Basicamente existem dois tipos de Modelos Multidimensionais na Teoria da Resposta ao Item, que são utilizados para respostas dicotômicas: Modelos compensatórios e modelos não-compensatórios. Um modelo é dito compensatório quando a mesma probabilidade de acerto do item é mantida quando a diminuição do valor de um traço latente é compensada pelo aumento no valor de outro traço latente.

Modelo de Ogiva Normal de três parâmetros

Os modelos de ogiva normal são muito utilizados e encontram-se implementados em diversos programas existentes, onde estimam-se os parâmetros pelos métodos dos mínimos quadrados, método da máxima verossimilhança, métodos bayesianos e moda

a posteriori.

$$P(X_i = 1|a_i, b_i, c_i, \theta) = c_i + (1 - c_i)\Phi\left(b_i + \sum_{k=1}^p a_{ik}\theta_k\right) = \Phi_i(\theta) \quad (1.18)$$

Onde $i = 1, 2, \dots, I$, sendo I o número de itens na prova, e $k = 1, 2, \dots, p$, sendo p a dimensão do vetor de traços latentes. O parâmetro b_i é denominado dificuldade do i -ésimo item e cada componente do vetor $a_i = (a_{1i}, \dots, a_{pi})$ é a discriminação do i -ésimo item sobre o k -ésimo traço latente. Quando as habilidades são baixas, incorpora-se uma probabilidade de acerto casual c_i , $i = 1, \dots, I$.

Modelo Logístico Compensatório de três parâmetros

McKinley e Reckase (1980) propõem o modelo logístico compensatório de três parâmetros dado pela expressão:

$$P(X_i = 1|a_i, b_i, c_i, \theta) = c_i + (1 - c_i)\frac{1}{1 + e^{(-\sum_{k=1}^p a_{ik}\theta_k + b_i)}} = P_i(\theta) \quad (1.19)$$

Onde os parâmetros são definidos como o modelo anterior.

Modelos não compensatórios

O Modelo logístico não compensatório de três parâmetros proposto por Whitely (1980) e Simpson (1977) é dado por:

$$P(X_i = 1|\theta, a_i, b_i, c_i) = c_i + (1 - c_i)\prod_{k=1}^p \frac{1}{1 + e^{(-Da_{ik}\theta_{jk} + b_{ik})}} \quad (1.20)$$

Onde a , c_i e θ são definidos como o modelo anterior e $b = (b_{i1}, \dots, b_{pi})$ representa a dificuldade do item em questão associada a cada traço latente.

Capítulo 2

Dados Faltantes

2.1 Introdução

Perda de informações é um problema frequente em diversas áreas do conhecimento científico. Pesquisadores se deparam com dados incompletos (isto é, conjuntos de dados que contém dados faltantes) por diferentes motivos. A não resposta pode ocorrer durante a realização da pesquisa ou avaliação, como a falta cooperação do entrevistado; questionários ou testes muito longos e cansativos; tempo curto para responder aos quesitos; falta de conhecimento para responder aos itens; questões mal formuladas, etc... Os dados ausentes também podem surgir na coleta dos dados (por exemplo, erro de digitação ao fazer o registro dos dados).

Com a exceção de McKnight et. al. (2007, capítulo 4), livros sobre dados faltantes não mencionam a prevenção. No entanto, prevenir é o ataque mais direto sobre os problemas gerados pela ausência de resposta. Alguns procedimentos podem melhorar a pesquisa, diminuindo a incidência de dados faltantes, são eles: minimizar o uso de medidas invasivas; usar incentivos para estimular a resposta; tentar igualar o entrevistador e entrevistado na mesma idade e etnia; adaptar o modo da pesquisa (por telefone, cara a cara, questionário web, e assim por diante) para a população em estudo; usar um *design* diferente para diferentes grupos em seu estudo; procurar os não respondentes e pedir que preencham os itens não respondidos; se possível, tentar recuperar os dados faltantes por outras fontes, etc...

Outros procedimentos podem ser tomados antes da aplicação da pesquisa ou

teste para evitar a não resposta: em estudos experimentais, tentar minimizar a carga e intensidade dos tratamentos sempre que possível; preparar um bom questionário, que explique a finalidade e utilidade do estudo; tentar organizar a coleta de dados por meio de uma autoridade, por exemplo, o próprio médico do paciente, no caso de pesquisas médicas; realizar um estudo piloto para detectar e atenuar os problemas, etc.

Procedimentos de ponderação: É um método para reduzir o viés quando a probabilidade de seleção no levantamento difere entre os entrevistados. Em pesquisas por amostragem, os respondentes são ponderados por pesos do *design*, que são inversamente proporcionais à sua probabilidade de serem selecionados na pesquisa. Se faltarem dados, os casos completos são reponderados de acordo com pesos do *design*, que são ajustados para combater os efeitos de seleção produzidos por não-resposta. O método é muito utilizado em estatísticas oficiais. Indicações importantes incluem Cochran (1977), Särndal et al. (1992) e Bethlehem (2002).

A presença de observações faltantes complica as análises estatísticas dos dados, pois muitos testes estatísticos são aplicados apenas em matriz de dados completos (isto é, banco de dados sem nenhuma perda de informação). Além disso, uma abordagem adequada para bancos de dados com observações incompletas é uma questão bastante delicada, pois a utilização de métodos inadequados pode levar a conclusões erradas sobre o conjunto de dados. Três tipos de problemas estão associados aos valores faltantes: perda de eficiência, complicações na manipulação e análise dos dados, e viés, resultante das discrepâncias entre os valores atribuídos aos dados faltantes e os valores reais desconhecidos (Farhangfar et al., 2007). O não tratamento ou o tratamento inadequado dos dados faltantes também pode afetar a generalização dos resultados da análise (Mcknight et al., 2007; Colantonio et al., 2010).

Para contornar esse problema, surgiram técnicas estatísticas que envolvem a substituição dos dados faltantes por estimativas de valores plausíveis a serem imputados aos dados faltantes. Estas técnicas têm por objetivo completar os bancos de dados e possibilitar a análise com todos os dados em estudo. As primeiras técnicas de imputação desenvolvidas envolviam métodos relativamente simples, tais como substituição dos dados faltantes pela média ou pela mediana da variável, por interpolação ou até por regressão linear. Entretanto existem desvantagens na utilização desses

métodos, como a subestimação da variabilidade e a impossibilidade da utilização de outras variáveis do próprio conjunto de dados para melhorar o processo de imputação.

Para definir o melhor método de imputação a ser utilizado é importante conhecer o mecanismo de não resposta, pois a partir dele pode se ter a idéia a respeito do relacionamento entre a perda da informação e os valores das variáveis presentes na matriz de dados.

2.1.1 Mecanismos de Dados Faltantes

Todos os trabalhos envolvendo complementação de dados faltantes levam inevitavelmente em conta o mecanismo que causou a ausência dos dados. Respostas aos itens podem ser classificadas como dados faltantes completamente aleatórios (MCAR), dados faltantes aleatórios (MAR), ou dados faltantes não aleatórios (MNAR).

Os dados faltantes completamente aleatórios (MCAR) ocorrem quando a probabilidade de um item ter respostas ausentes não depende de qualquer das quantidades, observadas ou não observadas. Ou seja, a probabilidade de ocorrência do dado ausente é a mesma para todos os casos. Isso significa efetivamente que as causas dos dados faltantes não estão relacionadas aos dados. Diferencia-se dos outros mecanismos, onde os valores ausentes não estão distribuídos aleatoriamente. Uma vez que esta pressuposição se mantém, o processo que gera os valores faltantes pode ser ignorado durante a análise e estimadores de uma análise de caso completo pode ser não tendenciosos (Horton & Kleinman, 2007; Molenberghs & Verbeke, 2005), mas podem ter substanciais perdas de eficiência (Little & Rubin, 2002).

No entanto, dados faltantes completamente aleatórios (MCAR) é uma hipótese forte e pouco realista, pois na maioria dos experimentos de pesquisa, quase sempre existe algum grau de relação entre os valores em falta e algumas das informações das co-variáveis. Assim, pode ser mais razoável afirmar que os dados são faltantes aleatórios (MAR), que ocorre quando um item faltante depende apenas de algumas quantidades observadas, que podem ser alguns resultados e/ou co-variáveis. Este mecanismo é considerado mecanismo de perda previsível, em que os dados são perdidos por um processo aleatório, quando a probabilidade de não resposta depende dos dados presentes (mas não dos ausentes).

Se os dados não podem ser considerados faltantes completamente aleatórios (MCAR) e nem faltantes aleatórios (MAR), então os itens faltantes dependem dos dados não observados e são ditos faltantes não aleatórios (MNAR). Por exemplo, respostas faltantes para um item podem ser devidas ao alto grau de dificuldade do item. Dados faltantes não aleatórios MNAR significa que a probabilidade de se ter um dado faltante varia por razões que são desconhecidas. Neste mecanismo, um valor ausente pode depender do próprio valor ausente.

As classificações de Rubin são importantes para entender por que alguns métodos não funcionam. Sua teoria estabelece as condições sob as quais um método de dados ausentes pode fornecer inferências estatísticas válidas. Correções mais simples só funcionam sob a suposição de MCAR, que é restritiva e muitas vezes irreal. Se MCAR é implausível, tais métodos podem fornecer estimativas tendenciosas.

Apesar dos mecanismos de dados ausentes precisarem ser levados em consideração durante a modelagem estatística, para produzir estimativas eficientes e não tendenciosas, segundo Molenberghs e Verbeke (2004), nenhuma modelagem abordada, seja por MAR ou por MNAR, pode compensar totalmente a perda de informação que ocorre, devido ao carácter incompleto dos dados.

O primeiro mecanismo (MCAR) possui uma suposição muito forte e raramente é satisfeito na prática. Quando ele ocorre, os dados não observados constituem uma sub amostra aleatória. O segundo mecanismo (MAR) também é denominado de Ignorável. O termo Ignorável é usado para indicar que não é necessário especificar um modelo para a não-resposta.

É importante ressaltar que o mecanismo de dados ausentes é que pode ser ignorado e não os dados ou as unidades com dados ausentes. O terceiro mecanismo de ausência é não-ignorável devido à falta de aleatoriedade da não-resposta. Portanto, nesta situação torna-se necessário especificar um modelo para a não resposta.

2.1.2 Métodos de Imputação de dados faltantes

Obviamente, a melhor forma de tratar dados faltantes é não tê-los (Orchard e Woodbury, 1972, p.697). Embora haja muita verdade nesta afirmação, esse ideal não é atingido na prática. O modo padrão de tratar dados faltantes é excluí-los.

Assim, métodos de exclusão, como *Listwise*, são utilizados e não são mencionados nas pesquisas. A presença de dados faltantes muitas vezes nem são explícitas no texto, gerando diferentes tabelas, baseadas em diferentes tamanhos de amostras. Mesmo os estudos mais cuidadosamente projetados e executados produzem valores faltantes. A questão que realmente interessa é a forma como lidamos com dados incompletos. Os tipos de imputação a seguir foram extraídos do livro *Flexible Imputation of Missing Data*, Van Buuren (2012):

Exclusão listwise:

Análise de caso completo listwise é o default para lidar com dados incompletos em muitos pacotes estatísticos, incluindo SPSS, SAS e Stata. A função `na.omit ()` faz o mesmo no S-PLUS e R. O procedimento elimina todos os casos com um ou mais valores ausentes nas variáveis de análise. A grande vantagem do processo de análise completa é a conveniência. Se os dados são MCAR, a exclusão listwise produz estimativas não viesadas sobre as médias, variância e os pesos da regressão. Sob MCAR, exclusão listwise produz erros padrão e níveis de significância que estão corretos para a redução do subconjunto de dados, mas que são muitas vezes maiores em relação a todos os dados disponíveis.

Uma desvantagem de exclusão listwise é que é, potencialmente, um desperdício. Não é raro que em aplicações reais mais de metade da amostra original seja perdida, especialmente se o número de variáveis é grande. King et al. (2001) estimaram que o percentual de registros incompletos em ciências políticas ultrapassou os 50% em média, havendo alguns estudos que têm mais de 90% de registros incompletos. É evidente que uma subamostra menor poderia degradar seriamente a capacidade para detectar os efeitos de interesse.

Além disso, se os dados não são MCAR, exclusão listwise pode severamente enviesar estimativas de médias, coeficientes de regressão e correlação. Little e Rubin (2002, pp 41-44) mostraram que o viés da média estimada aumenta com a diferença entre as médias dos casos observados e ausentes, e com a proporção dos dados ausentes. Schafer e Graham (2002) relataram um estudo de simulação que demonstra o viés da exclusão *listwise* sob MAR e MNAR. No entanto, as análises de caso completo nem sempre são ruins. As implicações dos dados ausentes são diferentes, dependendo de onde eles ocorrem (resultados ou preditores) e também depende dos parâmetros e

do modelo utilizado na análise dos dados completos. No contexto da análise de regressão, exclusão *listwise* possui algumas propriedades únicas que a torna atraente em ambientes particulares. Há casos em que exclusão *listwise* pode fornecer estimativas melhores do que até mesmo os procedimentos mais sofisticados.

A Exclusão *listwise* pode introduzir inconsistências no relatório final. Uma vez que essa exclusão é aplicada automaticamente para o conjunto ativo de variáveis, diferentes análises sobre os mesmos dados são muitas vezes baseadas em subamostras diferentes. A princípio, é possível produzir uma subamostra utilizando todas as variáveis globais ativas. Na prática, esta é pouco atrativa, a subamostra global terá sempre menos casos que cada uma das sub-amostras locais, por isso é comum criar subconjuntos diferentes para diferentes tabelas. É evidente que isto complica a comparação e generalização para a população em estudo.

Em alguns casos, a exclusão *listwise* pode levar a subamostras sem sentido. Seria muito mais difícil, se não impossível, fazer análises que envolvem o tempo, por exemplo, para identificar padrões semanais ou para ajustar modelos autoregressivos que predizem os dias anteriores.

Os autores líderes na área estão, no entanto, com medo de fornecer conselhos sobre a percentagem de casos ausentes abaixo da qual é aceitável fazer exclusões *listwise*. Little e Rubin (2002) argumentam que é difícil formular ‘regras de ouro’ uma vez que as consequências do uso de exclusões *listwise* dependem mais do que apenas da taxa de dados ausentes. Vach (1994, p 113) manifestou o seu desagrado pelas regras simplistas, como segue:

‘Costuma-se supor que existem algumas coisas, como a taxa crítica de dados ausentes, em que os valores ausentes não sejam muito danosos. A crença em dados faltantes com uma taxa tão global é bastante estúpida.’

Exclusões *Pairwise*:

As exclusões *Pairwise*, também conhecidas como análise de casos disponíveis, são tentativas para resolver o problema das perda de dados das exclusões *listwise*. O método calcula as médias e as (co) variâncias em todos os dados observados. Assim, a média da variável X é baseada em todos os casos com os dados observados em X, a média da variável Y utiliza todos os casos com valores de Y observados, e assim por diante. Para as correlações e covariância, todos os dados são tomados em que

ambos X e Y têm escores não ausentes. Posteriormente, a matriz das estatísticas sumárias é alimentada em um programa de análise de regressão, análise fatorial ou outros procedimentos de modelagem.

Os softwares SPSS, SAS e Stata contêm muitos procedimentos com opção para exclusão *pairwise*. O método é simples, utiliza todas as informações disponíveis e produz estimativas consistentes de médias, correlações e covariâncias sob MCAR (Littler e Rubin, 2002, p.55). No entanto, quando tomadas em conjunto, estas estimativas têm grandes deficiências. As estimativas podem ser tendenciosas, se os dados não são MCAR. Além disso, existem problemas computacionais. A matriz de correlação pode não ser positiva definida, que é requisito para a maioria dos procedimentos multivariados. Correlações fora do intervalo $[-1, 1]$ podem ocorrer, um problema que surge a partir de diferentes subconjuntos utilizados para as variâncias e covariâncias. Tais problemas são mais graves para as variáveis altamente correlacionadas (Little, 1992). Outro problema é que não é claro qual o tamanho da amostra deve ser usado para calcular os erros padrões. Tomando o tamanho médio de amostra, produz-se erros padrões que são demasiado pequenos (Little, 1992).

A idéia por trás da imputação *pairwise* é utilizar toda a informação disponível. Embora esta idéia seja boa, a análise adequada da matriz *pairwise* requer técnicas de otimização sofisticadas e fórmulas especiais para calcular o erro padrão (Van Praag et al, 1985; Marsh, 1998).

Imputação pela média

Imputação pela média é uma solução rápida e simples para os dados em falta. No entanto, podem ocorrer os seguintes problemas: subestimação da variância; perturbação nas relações entre as variáveis; viés em quase qualquer estimativa diferente da média e viés na estimativa da média quando os dados não são MCAR. A imputação pela média só deverá ser utilizada como uma correção rápida quando um pequeno número de valores estão faltando, e deve ser evitada em geral.

Imputação via Regressão

Imputação via regressão incorpora conhecimento de outras variáveis com a idéia de produzir imputações mais inteligentes. A primeira etapa envolve a construção de um modelo a partir dos dados observados. Previsões para os casos incompletos são, então, calculadas de acordo com o modelo ajustado, e servem como substitutos para

os dados faltantes.

A imputação via regressão produz estimativas não viesadas das médias sob MCAR, assim como a imputação pela média, e dos pesos de regressão do modelo de imputação, se as variáveis explicativas são completas. Além disso, os pesos de regressão são não viesados sob MAR, se os fatores que influenciam a falta completa fazem parte do modelo de regressão. Por outro lado, a variabilidade dos dados imputados é sistematicamente subestimada. O grau de subestimação depende da variância explicada e da proporção de casos em falta (Littler e Rubin, 2002, p. 64).

Imputar valores previstos pode render imputações realistas se a previsão é perto da perfeição. Se assim for, o método reconstrói os dados faltantes a partir dos dados disponíveis. Na maioria das aplicações, é improvável surgir este tipo de dados faltantes.

Imputação via regressão estocástica

Imputação via regressão estocástica é um refinamento da imputação pela regressão, pois acrescenta ruído para as previsões. Isso terá um efeito descendente sobre as correlações.

Nota-se que algumas novas complexidades surgem. Existem várias imputações com valores negativos. Tais valores podem ser não plausíveis. A Imputação via regressão estocástica é um importante passo. Em particular, não só mantém a regressão, mas também a correlação entre as variáveis. Imputação via regressão estocástica não resolve todos os problemas e constitui a base de técnicas de imputação mais avançadas.

Imputação via última observação realizada (LOCF) e observação da linha de base transitadas (BOCF)

A Última observação realizada (LOCF) e a observação da linha de base transitadas (BOCF) requerem dados longitudinais. A idéia é levar o último valor observado como um substituto para os dados em falta.

LOCF é conveniente porque gera um conjunto de dados completo. O método é utilizado em ensaios clínicos. Nos EUA, Food and Drug Administration (FDA) tem tradicionalmente visto LOCF como o método preferido de análise, considerando-o conservador e menos propenso a seleção de exclusão listwise. No entanto, Molen-

berghs e Kenward (2007 , pp 47-50) mostram que o viés pode operar em ambos os sentidos , e que LOCF pode render estimativas tendenciosas, mesmo sob MCAR . LOCF deve ser seguido por um método de análise estatística adequado, que distingue entre os dados reais e imputados. Isto geralmente não é feito , no entanto.

Método do indicador

Suponha que se quer utilizar uma regressão, mas faltam valores em uma das variáveis explicativas. O método do indicador (Miettien, 1985, p. 232) substitui cada valor em falta por um zero e prolonga-se o modelo de regressão pelo indicador de resposta. O procedimento é aplicado a cada variável incompleta.

Este método é popular em saúde pública e epidemiologia. Uma vantagem é que o método do indicador conserva o conjunto de dados completo. Além disso, ele permite as diferenças sistemáticas entre os dados observados e não observados, pela inclusão do indicador de resposta. No entanto, o método pode produzir estimativas de regressão severamente viesadas, mesmo sob MCAR e para baixas quantidades de dados em falta (Vâch e Blettner, 1991; Greenland e Finkle, 1995;. Knol et al, 2010).

Em contrapartida, White e Thompson (2005) apontam que o método pode ser útil para estimar o efeito do tratamento em ensaios clínicos randomizados, quando a covariável base é parcialmente observada. Se os dados faltantes são restritos à covariável, se o interesse é apenas restrito à estimativa do efeito do tratamento, se a aderência ao tratamento alocado é perfeita e se o modelo é linear, sem interações, então utiliza-se o método do indicador, para que a covariável produza um estimativa não tendenciosa do efeito do tratamento. Isto é verdade mesmo que a falta completa dependa da própria covariância. Estas condições são muitas vezes difíceis de conseguir na prática. O método não permite a falta de dados nos resultados, os quais ocorrem com frequência em dados reais. Enquanto o método do indicador pode ser conveniente em alguns casos especiais, fica aquém, de um modo geral, para tratar os dados faltantes.

A Imputação via regressão e a imputação via regressão estocástica podem produzir estimativas imparciais em MAR. Para funcionar, o modelo precisa ser corretamente especificado. LOCF e o método do indicador são incapazes de fornecer estimativas consistentes, mesmo sob MCAR.

Exclusão listwise produz erros padrão que são corretos para o subconjunto de casos completos, mas, em geral, muito grande para o conjunto de dados reais. O cálculo de Erros padrão em exclusão pairwise é complicado. Os erros padrões após imputação são muito pequenos já que os cálculos padrões não fazem distinção entre os dados observados e os dados imputados. Fatores de correção para algumas situações têm sido desenvolvidos (Shafer e Schenker, 2000), mas uma solução mais conveniente é imputação múltipla.

Imputação Múltipla:

A Imputação múltipla cria $m > 1$ conjuntos de dados completos. Cada um destes conjuntos de dados são analisados por software em análises padrões. Os m resultados são agrupados em uma estimativa pontual final acrescidos do desvio padrão, por regras de agrupamento simples ("regras de Rubin"). Existem três etapas principais na imputação múltipla: imputação, análise e agrupamento.

A análise começa com dados observados, dados incompletos. A imputação múltipla cria várias versões completas dos dados, substituindo os valores em falta por valores de dados plausíveis. Estes valores plausíveis são extraídos de uma distribuição modelada especificamente para cada entrada de dados faltantes. Os conjuntos de dados imputados são idênticos para as entradas de dados observados, mas diferem nos valores imputados. A magnitude destas diferenças reflete nossa incerteza sobre o valor a imputar.

O segundo passo é estimar os parâmetros de interesse de cada conjunto de dados imputados. Isso geralmente é feito através da aplicação do método de análise que teríamos usado se os dados fossem completos. Os resultados serão diferentes porque os seus dados de entrada são diferentes. É importante perceber que essas diferenças são causadas apenas por causa da incerteza sobre o valor a imputar.

O último passo é a associação dos m parâmetros estimados em uma estimativa, e a estimação da variância. A variância combina a variância de amostragem convencional (variância intra-imputação) e a variância adicional causada pela falta de dados (variância entre imputação). Sob as condições adequadas, as estimativas combinadas são não viesadas e têm as propriedades estatísticas corretas.

A Imputação múltipla (Rubin, 1987a, 1996) resolve o problema de erros

padrões muito pequenos. Imputação múltipla é única no sentido de que ela fornece um mecanismo para lidar com a incerteza inerente das próprias imputações.

O nível de confiança em um valor imputado particular é expresso como a variação entre os m conjuntos de dados completos. Por exemplo, em uma pesquisa sobre deficiência, suponha que o entrevistado respondeu o item se ele podia andar, mas suponha que o entrevistado respondeu o item se ele poderia se levantar de uma cadeira. Se a pessoa pode andar, então é altamente provável que a pessoa também vai ser capaz de se levantar da cadeira. Assim, para as pessoas que podem andar, podemos tirar um "sim" para a falta de "se levantar de uma cadeira", com uma alta probabilidade, digamos 0,99, e usar o valor como o valor imputado. Em casos extremos, se estamos realmente certos, sempre iremos atribuir o mesmo valor para essa pessoa. De modo mais geral, que estão menos confiantes sobre o verdadeiro valor. Suponha-se que, em um estudo de crescimento, altura está ausente para um sujeito. Se nós só sabemos que essa pessoa é uma mulher, isso fornece algumas informações sobre os valores possíveis, mas nem tanto. Assim, o intervalo de valores plausíveis é muito maior aqui. A imputação é capaz de lidar com ambas as situações de alta confiança e baixa confiança igualmente bem.

Imputação de dados faltantes para dados categóricos

A Imputação de dados faltantes categóricos é possível sob uma ampla classe de modelos lineares generalizados (McCullagh e Nelder, 1989). Para variáveis binárias incompletas usa-se a regressão logística. Como a regressão logística será utilizada como proposta

2.1.3 Dados Faltantes em Teoria de Resposta ao Item

Em pesquisas educacionais, dados faltantes são sempre presentes em respostas a itens de um teste. Pesquisadores mostraram que ignorar tais dados faltantes pode criar problemas na estimação de parâmetros. Consequentemente, surgiram vários métodos de imputação de dados faltantes e vários estudos investigaram a tendenciosidade das estimativas dos parâmetros na TRI.

O uso da TRI combina as dificuldades do item com a proficiência das pessoas (Wauters, Desmet, e Van den Noortgate, 2010). Um número elevado de repostas faltantes apresenta problemas para utilização da TRI. Por exemplo, os problemas de não convergência na estimativa dos parâmetros do modelo da TRI, ao estimar as dificuldades de itens e as habilidades da pessoa. Além disso, durante a análise estatística dos dados ausentes, mecanismos que conduzem a respostas com itens ausentes precisam ser identificados, a fim de evitar estimativas tendenciosas (Little e Rubin, 2002).

Inicialmente, nas equações de verossimilhança para a curva característica do item (ICC) descrita (Lord, 1953), parecia haver três grandes obstáculos nas aplicações práticas:

- Soluções das equações de verossimilhança para os dados reais pareciam impraticáveis de um ponto de vista computacional (Torgerson, 1958, p.388);
- A teoria não leva em consideração que alguns examinados são impedidos de finalizar o teste, por causa da limitação do tempo disponível para responder aos quesitos (*speededness*);
- A teoria foi inicialmente desenvolvida para itens dicotômicos, enquanto que na verdade existem três tipos distintos de resposta: resposta correta, resposta incorreta e ausência de resposta (dados faltantes). Diversas investigações concluíram que respostas omitidas não devem ser tratadas como erradas nas equações de verossimilhança.

Mais tarde, Lord (1968), Bock & Lieberman (1970) e Bock (1972) conseguiram resolver as equações numéricas para as equações de verossimilhança com sucesso. Até mesmo a estimativa do parâmetro que representa a assíntota inferior da CCI (chamado de probabilidade de acerto ao acaso) pôde ser obtida por máxima verossimilhança (Wingersky & Lord, 1973), desde que haja dados suficientes para determinar esta parte da curva. A partir daí, muitos estudos tentaram resolver os problemas de dados faltantes e *speededness*.

Speededness:

Se um examinando não alcança um item por causa do limite de tempo, este fato não contém nenhuma informação útil para inferir sobre o seu nível de habilidade θ da curva característica do item (CCI). O modelo unidimensional CCI não fornece nenhuma maneira de fazer uso de qualquer relacionamento (normalmente curvilíneo) que possa existir entre a velocidade de resposta e θ .

Um teste com limite de tempo, para alguns examinados, está medindo a velocidade, diferente da medida de habilidade que seria obtida se o teste fosse administrado sem limite de tempo. A teoria de Resposta ao item não está preparada para lidar com a dimensão da velocidade, mas pode ser usada para analisar as respostas obtidas nas administrações de testes cronometrados. Para isso ser possível é necessário que os examinados respondam os itens seguindo a sequência, do primeiro até o último item que é possível responder naquele intervalo de tempo predeterminado. Para cada examinando, os itens seguintes a sua última resposta registrada (este item é chamado o último item tentado) são ignorados durante o processo de estimativa. Na prática, pode-se eliminar das análises examinados que respondam menos de um terço, por exemplo, dos n itens. Assim, a habilidade θ do examinado é estimada a partir de suas respostas aos itens presumidamente alcançados e os parâmetros dos itens são calculados a partir das respostas dos examinados que se presume ter atingido os itens. Isso não complica as equações de verossimilhança ou o processo de resolvê-las. Uma propriedade fundamental da teoria da TRI é que os parâmetros dos itens não dependem do grupo de examinados testado, dentro de limites razoáveis, e que o examinado com habilidade θ não depende dos itens administrados, assumindo que todos os itens medem a mesma dimensão psicométrica. Assim, ignorar vários examinados e vários itens não deve ter efeitos graves. É certo que o θ estimado desta forma aproxima a habilidade do examinado apenas se as respostas dele aos itens realmente tentados teriam sido o mesmo na falta de um limite de tempo. Independentemente disso, o θ estimado desta forma reflete adequadamente níveis de habilidade eficazes do examinando, se as condições de cronometragem forem realmente prestadas.

Capítulo 3

Estudos de Simulação

3.1 Introdução

Dados faltantes prejudicam as análises e podem ocorrer devido a vários mecanismos e fatores. Por exemplo, pessoas podem deixar em branco itens que se mostraram muito difíceis, podem perder o interesse no meio do teste, pular algumas seções, inadvertidamente, ou podem recusar-se a responder assuntos delicados. Uma grande quantidade de dados faltantes também podem surgir quando os respondentes navegam livremente nos itens e só respondem um número relativamente pequeno de exercícios de um vasto número que normalmente são fornecidos dentro de ambientes de aprendizagem (Wauters et al. 2010).

Enquanto não se sabe os verdadeiros motivos para a presença de dados faltantes, é comum ignorá-los ou tratá-los como respostas incorretas. Ambas as abordagens parecem ser problemáticas, uma vez que a primeira ignora todos os dados faltantes e a segunda assume, com certeza, que a resposta correta é desconhecida, independente da habilidade do entrevistado. Em alguns estudos as respostas ausentes são tratadas de maneiras diferentes em diferentes fases das análises. Por exemplo, ao utilizar a TRI, é comum tratar os dados faltantes como ignoráveis, na fase de estimação dos parâmetros dos itens, e, posteriormente, tratá-los como resposta incorreta, na fase da estimação da proficiência.

Nesta dissertação, analisaremos três efeitos de tratamentos de dados faltantes nas estimativas dos parâmetros da TRI unidimensional de três parâmetros:

1. Ignorar os dados faltantes;
2. Assumir que os dados faltantes são respostas incorretas;
3. Em duas fases: ignorar os dados faltantes no processo de calibração e considerá-los como resposta incorreta no processo de estimação das proficiências.

Teorias, metodologias e softwares para lidar com problemas de dados incompletos foram vastamente ampliados e aperfeiçoados ao longo das últimas décadas. Os principais pacotes de análise estatística, atualmente, possuem procedimentos para a realização das análises apropriadas. Nesta dissertação, são sugeridas simulações para avaliar o efeito de dados faltantes não ignoráveis na estimação dos parâmetros dos itens e da proficiência, utilizando modelos da TRI.

Para o exemplo simulado, o ajuste do modelo e a estimação dos parâmetros foram realizados utilizando as funções “tpm”, “eap”, “coef” e “factor.scores” da biblioteca “irtoys” do software R (Andrade & Anjos, 2012).

Neste capítulo, foram simulados quatro casos:

1. Simulação sem dados faltantes
2. Simulação utilizando dados faltantes ignorados
3. Simulação com dados faltantes como respostas incorretas
4. Simulação considerando os dados faltantes: como ignorados, no processo de calibração e como respostas incorretas, nas estimativas das proficiências.

No próximo capítulo, serão propostos dois modelos para imputação de dados faltantes, em que a ausência de resposta é tratada como não ignorável:

1. Imputação de dados via regressão logística;
2. Imputação de dados via regressão isotônica suavizada;

3.1.1 Metodologia para simulação

Para simulação foram utilizados os parâmetros gerados (valores reais dos parâmetros) e os parâmetros estimados.

Os parâmetros gerados (ou parâmetros reais) foram gerados aleatoriamente da seguinte forma:

1. Parâmetro de discriminação do item gerado: 50 valores gerados aleatoriamente a partir de uma distribuição uniforme no intervalo $[0.2, 3]$.
2. Parâmetro de dificuldade do item gerado: 50 valores distribuídos de forma equidistante no intervalo $[-2, 2]$.
3. Valor do parâmetro de acerto ao acaso: igual a 0.2 para os 50 itens.
4. Proficiência gerada : 1000 valores gerados através de uma distribuição $N(0,1)$.

Utilizando os valores dos parâmetros gerados (valores reais dos parâmetros), foram simuladas duas matrizes: uma matriz de dados completos e uma matriz de dados incompletos, da seguinte forma:

Foram geradas respostas binárias (certo=1, errado=0) utilizando o modelo de TRI de três parâmetros, com 50 colunas (itens) e 1000 linhas (respondentes), utilizando um valor (semente), que possa ser usado para produzir a mesma matriz em outras simulações. Esta matriz é completa (não contém dados faltantes). Utilizando-se a matriz de ‘zeros’ e ‘uns’ simulada, foram estimados, através do modelo de TRI de três parâmetros:

1. Parâmetro de discriminação do item estimado: 50 valores foram estimados.
2. Parâmetro de dificuldade do item estimado: 50 valores foram estimados
3. Proficiência estimada: 1000 valores foram estimados.
4. Neste capítulo, o parâmetro de acerto ao acaso (c) foi mantido fixo com valor igual a 0.2, não sendo estimado. No próximo capítulo, para as propostas de imputação de dados, os 50 valores do parâmetro de acerto ao acaso foram estimados.

A matriz incompleta foi construída com a retirada de cerca de 25 % de dados da matriz completa de ‘zeros’ e ‘uns’ simulada, da seguinte forma: os dados faltantes

foram gerados aleatoriamente sendo que a probabilidade de não resposta foi diretamente proporcional à probabilidade do respondente não saber a resposta para o item, dada por:

$$1 - \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad (3.1)$$

Esta probabilidade foi multiplicada por um fator r no intervalo $[0,1]$. Considerando $r = 0.5$, obtivemos cerca de 25 % de dados faltantes. Este procedimento faz com que os dados faltantes não se caracterizem como “dados faltantes completamente ao acaso”.

As simulações foram realizadas considerando os parâmetros gerados (valores reais dos parâmetros) e os estimados e, também, a matriz completa e a incompleta.

3.1.2 Caso 1: Simulação sem dados faltantes

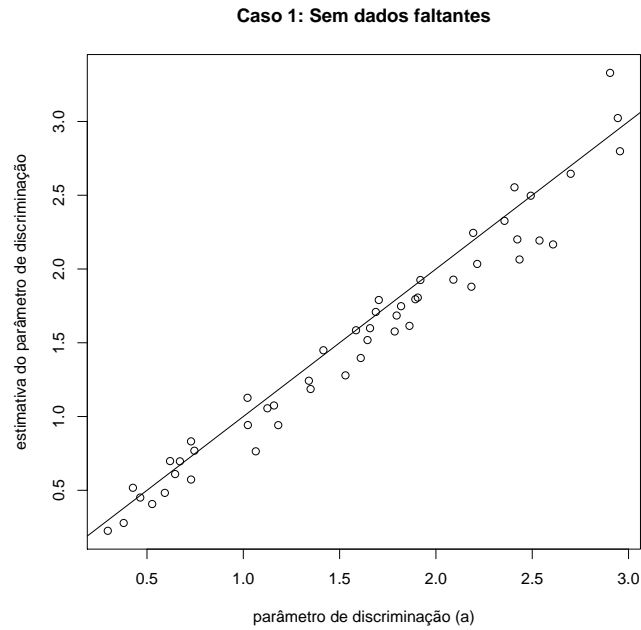


Figura 3.1: Caso 1: Simulação sem dados faltantes. Diagrama de dispersão entre o parâmetro de discriminação do item gerado (valores reais dos parâmetros) e o parâmetro de discriminação do item estimado

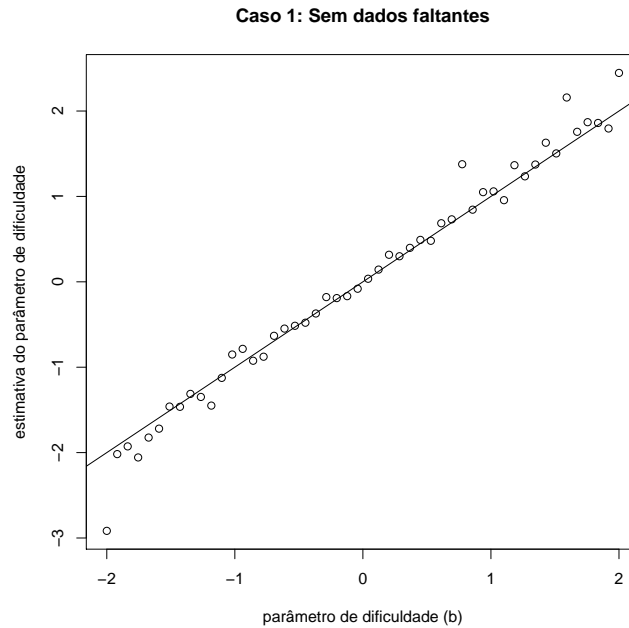


Figura 3.2: Caso 1: Simulação sem dados faltantes. Diagrama de dispersão entre o parâmetro de dificuldade do item gerado (valores reais dos parâmetros) e o parâmetro de dificuldade do item estimado

O caso 1 trata da simulação de dados a partir de um modelo de TRI unidimensional, usando a matriz de dados completos (isto é, não há nenhuma resposta faltante) e será utilizado para fornecer um padrão para comparações com os outros casos analisados.

Para o caso 1 foram utilizados os parâmetros gerados e os parâmetros estimados a partir da matriz completa de dados.

Os valores de discriminação do item, de dificuldade do item e de proficiência, estimados a partir da matriz completa de ‘zeros’ e ‘uns’ simulada (matriz de dados completos), foram ‘plotados’ no eixo vertical dos 3 gráficos correspondentes às figuras 3.1, 3.2 e 3.3, respectivamente. No eixo ‘horizontal’ foram plotados os valores que foram gerados aleatoriamente (valores reais dos parâmetros). Uma linha de tendência linear foi introduzida em cada um dos três gráficos, para servir de indicação do grau de igualdade entre valores nos dois eixos.

A figura 3.1 apresenta, no eixo vertical, os 50 valores dos parâmetros de discriminação do item estimados. Os valores foram estimados através da matriz de

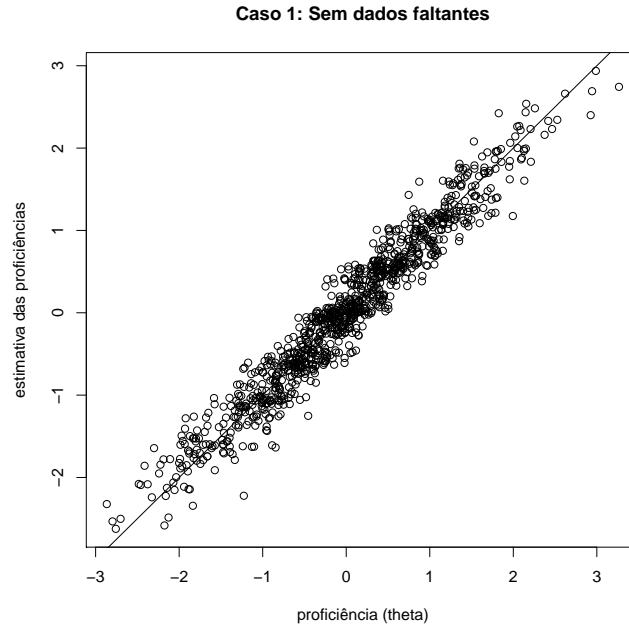


Figura 3.3: Caso 1: Simulação sem dados faltantes. Diagrama de dispersão entre a proficiência gerada e a proficiência estimada

‘zeros’ e ‘uns’ (dados completos), utilizando o modelo de TRI para três parâmetros. No eixo horizontal estão os 50 parâmetros de discriminação do item gerado, que foram gerados aleatoriamente a partir de uma distribuição uniforme no intervalo $[0.2, 3]$. Nota-se uma leve tendência de subestimação dos valores dos parâmetros de discriminação.

A figura 3.2 apresenta, no eixo vertical, os 50 valores dos parâmetros de dificuldade do item estimados através da matriz de ‘zeros’ e ‘uns’, utilizando o modelo de TRI para três parâmetros. No eixo horizontal estão os 50 valores dos parâmetros de dificuldade do item gerado, distribuídos de forma equidistante no intervalo $[-2, 2]$. Observa-se que as estimativas de dificuldade dos itens estão próximos dos valores reais gerados para os parâmetros.

A figura 3.3 apresenta, no eixo vertical, os 1000 valores de proficiência estimada. Os valores foram estimados através da matriz de ‘zeros’ e ‘uns’, utilizando o modelo de TRI para três parâmetros. No eixo horizontal, estão os 1000 valores de proficiência gerada (valores reais de proficiências individuais). Os valores foram gerados através de uma distribuição $N(0,1)$. Observa-se, também, que os valores estimados

distribuem-se em torno dos valores reais.

Conclui-se que, para o caso 1, todas os parâmetros estimados estão bem ajustados aos parâmetros gerados (valores reais dos parâmetros), exceto para o parâmetro de discriminação, que estão levemente subestimados.

3.1.3 Caso 2: Simulação utilizando dados faltantes ignorados

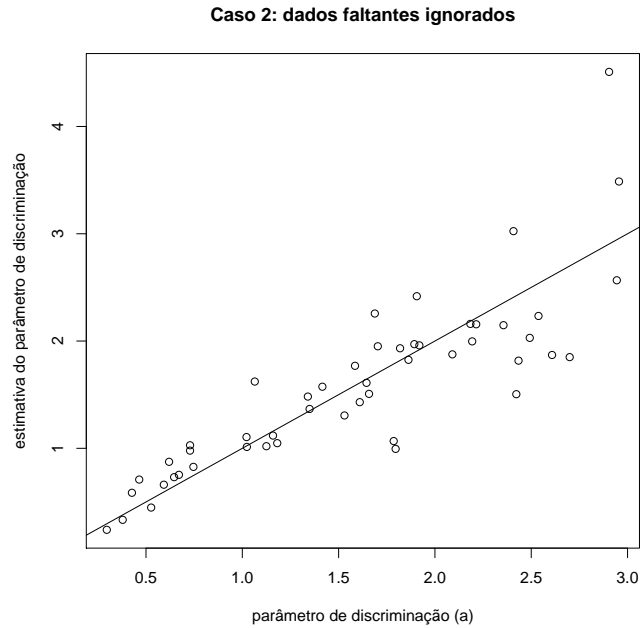


Figura 3.4: Caso 2: Simulação com dados faltantes ignorados. Diagrama de dispersão entre o parâmetro de discriminação do item gerado (valores reais dos parâmetros) e o parâmetro de discriminação do item estimado

O caso 2 considera o modelo TRI unidimensional onde os dados faltantes são tratados como ignorados. As especificações do modelo são as mesmas utilizadas no caso 1.

As figuras 3.4, 3.5 e 3.6 apresentam, nos eixos horizontais, os mesmos valores gerados aleatoriamente no caso 1: 50 parâmetros de discriminação do item, 50 parâmetros de dificuldade do item 1000 valores de proficiências. No eixo vertical de cada gráfico estão os parâmetros de discriminação e de dificuldade dos itens e as proficiências, estimados a partir de um modelo de TRI de três parâmetros, utilizando

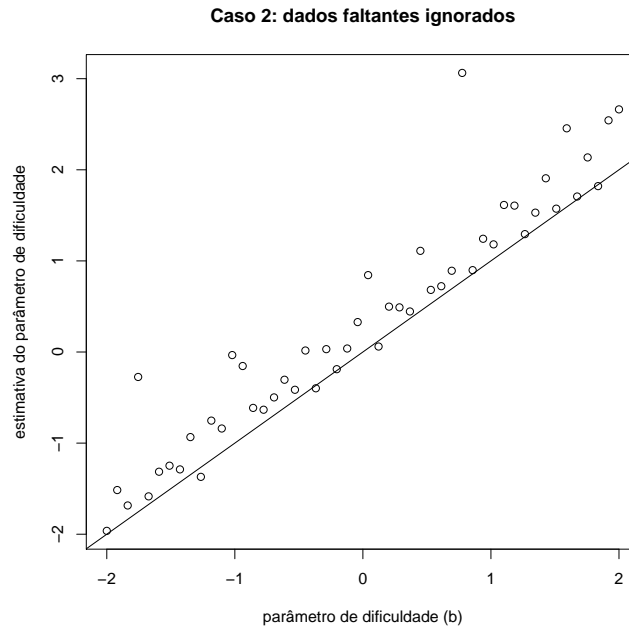


Figura 3.5: Caso 2: Simulação com dados faltantes ignorados. Diagrama de dispersão entre o parâmetro de dificuldade do item gerado (valores reais dos parâmetros) e o parâmetro de dificuldade do item estimado

uma matriz incompleta, formada pela retirada aleatoria de cerca de 25 % de dados da matriz completa de ‘zeros’ e ‘uns’, a mesma gerada no caso 1, conforme a metodologia apresentado anteriormente.

De acordo com a figura 3.4, o tratamento dos dados faltantes como ignorados tendem a superestimar levemente os parâmetros de discriminação a de menor valor. Verifica-se uma maior dispersão das estimativas para valores maiores de a .

Na figura 3.5, verifica-se que praticamente todos os valores estimados dos parâmetros de dificuldades ficaram superestimados em relação aos valores reais dos parâmetros.

Os valores das proficiências da figura 3.6 não apresentam a mesma dispersão dos valores encontrados no caso 1, pois estão menos concentrados em torno da linha de tendência, principalmente nos casos de proficiências com valores mais baixos, onde nota-se uma tendência a superestimação.

Conclui-se que tratar os dados faltantes tratados como ignorados causa viés nas estimativas dos parâmetros dos itens e das proficiências.

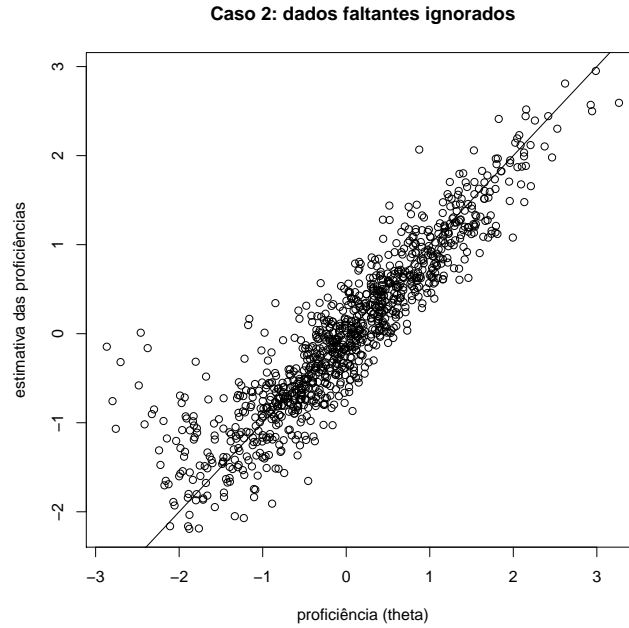


Figura 3.6: Caso 2: Simulação com dados faltantes ignorados. Diagrama de dispersão entre a proficiência gerada (valores reais dos parâmetros) e a proficiência estimada

3.1.4 Caso 3: Simulação com dados faltantes como respostas incorretas

No caso 3 o modelo TRI unidimensional é ajustado usando dados onde respostas omitidas foram tratadas como incorretas, tendo, assim, uma matriz de dados completa. A especificação de modelo é equivalente aos casos 1 e 2, mas as estimativas baseiam-se em dados recodificados, onde os dados faltantes da mesma matriz incompleta do modelo 2 são preenchidos por ‘zeros’.

As figuras 3.7, 3.8 e 3.9 apresentam, nos eixos horizontais, os mesmos valores utilizados nos casos 1 e 2, ou seja, os 50 valores de discriminação do item, 50 valores de dificuldade do item e os 1000 valores das proficiências, gerados aleatoriamente através de uma distribuição uniforme $[0.2, 3]$, de forma equidistante no intervalo $[-2, 2]$ e através de uma distribuição normal $N(0, 1)$, respectivamente, conforme a metodologia apresentada.

No eixo vertical dos gráficos 3.7, 3.8 e 3.9 estão as 50 estimativas de discriminação do itens, as 50 estimativas de dificuldade dos itens e as 1000 estimativas de

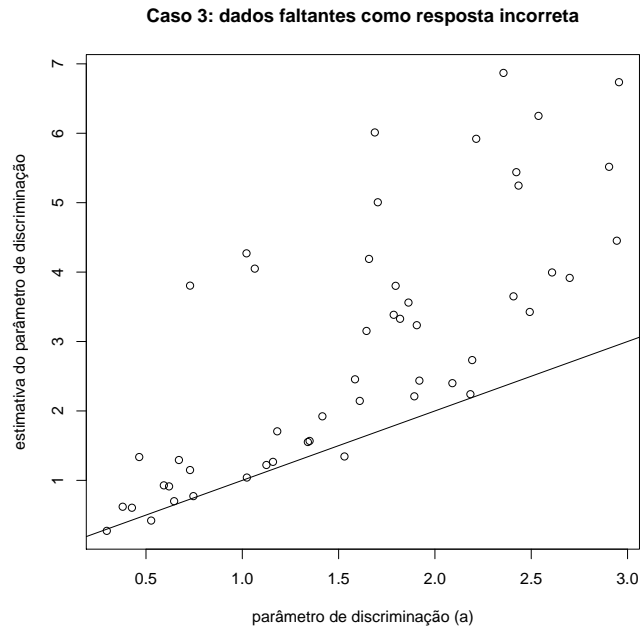


Figura 3.7: Caso 3: Simulação com dados faltantes como respostas incorretas. Diagrama de dispersão entre o parâmetro de discriminação do item gerado (valores reais dos parâmetros) e o parâmetro de discriminação do item estimado

proficiência, respectivamente. Estas estimativas foram feitas através de utilização da TRI de três parâmetros, sobre os mesmos valores da matriz de ‘zeros’ e ‘uns’ do caso 1, com 50 colunas (itens) e 1000 linhas (respondentes), sendo que os mesmos 25% dos dados foram apagados, da mesma forma que o caso 2, e foram recodificados como respostas incorretas (zeros).

Na figura 3.7, observa-se que as estimativas dos parâmetros de discriminação a dos itens estão mais superestimados do que nos casos 1 e 2 apresentados. Assim como no caso 2, os menores valores de a estão mais concentrados em torno da linha de tendência do que os valores maiores, para os quais estimativas apresentam maior variabilidade.

Assim como no caso 2, a figura 3.8 apresenta praticamente todos os parâmetros de dificuldade dos itens superestimados, ao utilizar dados faltantes como respostas incorretas.

Já os valores estimados das proficiências, na figura 3.9, não ficaram prejudicados como no caso 2, pois estão concentrados em torno da linha de tendência como

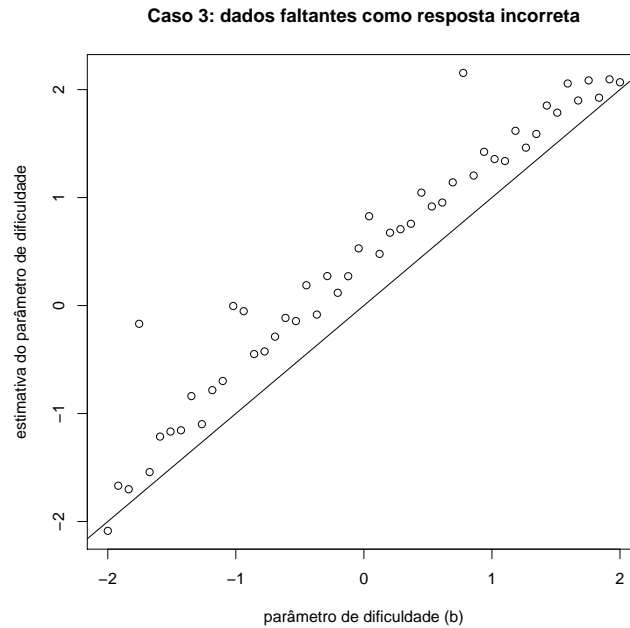


Figura 3.8: Caso 3: Simulação com dados faltantes como respostas incorretas. Diagrama de dispersão entre o parâmetro de dificuldade do item gerado e o parâmetro de dificuldade do item estimado

no caso 1.

Conclui-se que, tratar dados faltantes como respostas erradas causa viés das estimativas dos parâmetros de dificuldade e de discriminação dos itens. Porém, as estimativas dos valores da proficiência ficam próximas dos valores reais (proficiência gerada).

3.1.5 Caso 4: Simulação ignorando dados faltantes no processo de calibração e estimativas das proficiências considerando os dados faltantes como respostas incorretas.

O caso 4 também tem as mesmas especificações dos casos 1 a 3. Os parâmetros dos itens foram tomados como fixos, com os mesmos valores do caso 2. Este modelo reproduz os procedimentos utilizados em muitas avaliações que utilizam Teoria de Resposta ao Item. Ou seja, na fase de calibração dos itens os dados faltantes são ignorados (são estimados como no caso 2). Na fase posterior, onde os parâmetros da

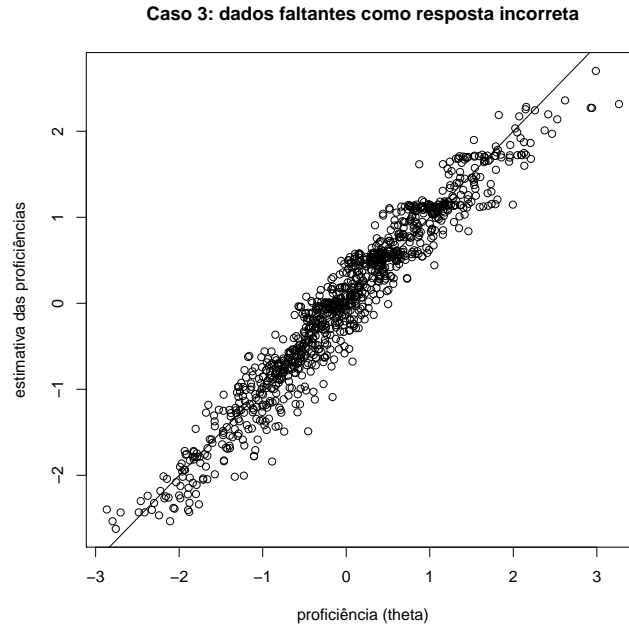


Figura 3.9: Caso 3: Simulação com dados faltantes como respostas incorretas. Diagrama de dispersão entre a proficiência gerada (valores reais dos parâmetros) e a proficiência estimada

pessoa e sua distribuição são estimados, considerando os parâmetros do item fixados através da calibragem de item, os dados faltantes são tratados como incorretos (como no caso 3).

As figuras 3.10, 3.11 e 3.12 apresentam, nos eixos horizontais, os mesmos valores utilizados nos casos 1 e 2 e 3, ou seja, os 50 valores de discriminação do item, 50 valores de dificuldade do item e os 1000 valores das proficiências, gerados aleatoriamente através de uma distribuição uniforme $[0.2, 3]$; através de uma forma equidistante no intervalo $[-2, 2]$ e através de uma distribuição normal $N(0, 1)$, respectivamente.

No eixo vertical das figuras 3.10, 3.11 e 3.12 estão as 50 estimativas de discriminação do itens, as 50 estimativas de dificuldade dos itens e as 1000 estimativas de proficiência, respectivamente. Estas estimativas foram feitas através da utilização da TRI de três parâmetros, sobre os mesmos valores da matriz de ‘zeros’ e ‘uns’ do caso 1, com 50 colunas (itens) e 1000 linhas (respondentes), sendo que os mesmos 25% dos dados foram apagados, da mesma forma que o caso 2. As estimativas dos parâmetros

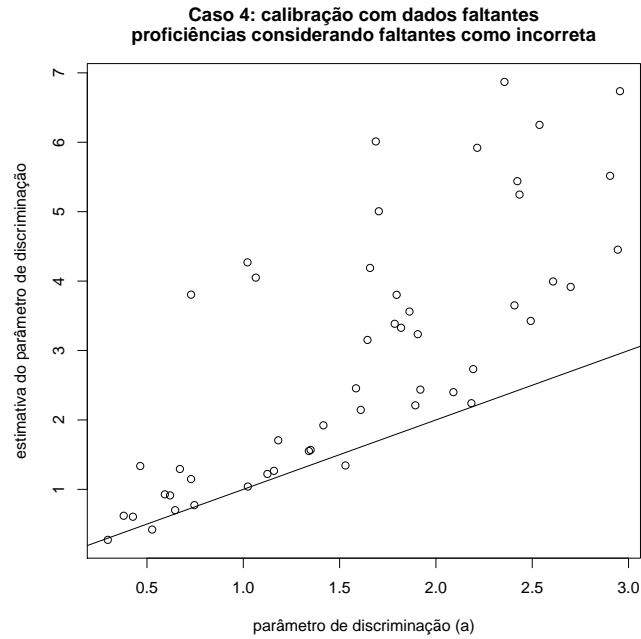


Figura 3.10: Caso 4: Simulação com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências. Diagrama de dispersão entre o parâmetro de discriminação do item gerado (valores reais dos parâmetros) e o parâmetro de discriminação do item estimado

dos itens foram realizados considerando os dados faltantes como ignorados (como no caso 2) e são fixadas. Após a fixação dos parâmetros dos itens, os dados faltantes são recodificados como respostas incorretas e, então, os valores da proficiência são estimados.

Na figura 3.10, observa-se que os parâmetros estimados de discriminação dos itens estão distribuídos como no caso 3, ou seja, as estimativas dos parâmetros de discriminação do item estão mais superestimadas do que nos casos 1 e 2.

Na figura 3.11, observa-se que a estimação dos parâmetros de dificuldade dos itens apresentam tendência a subestimação.

Na figura 3.12, observa-se que os valores de proficiência estimados estão mais dispersos que no caso 1, sendo que a dispersão é maior para menores valores de θ , apresentando tendência de superestimação para estes valores.

Assim, conclui-se que, para o caso 4, as estimativas dos parâmetros apresentam-se viesadas.

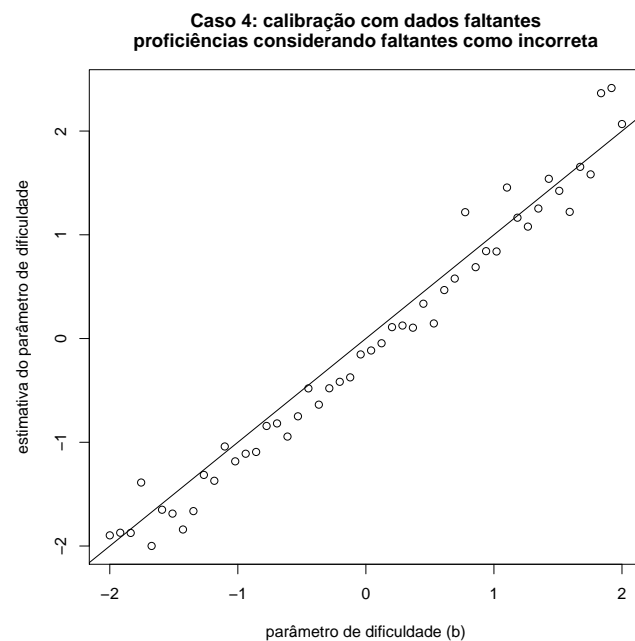


Figura 3.11: Caso 4: Simulação com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências. Diagrama de dispersão entre o parâmetro de dificuldade do item gerado (valores reais dos parâmetros) e o parâmetro de dificuldade do item estimado

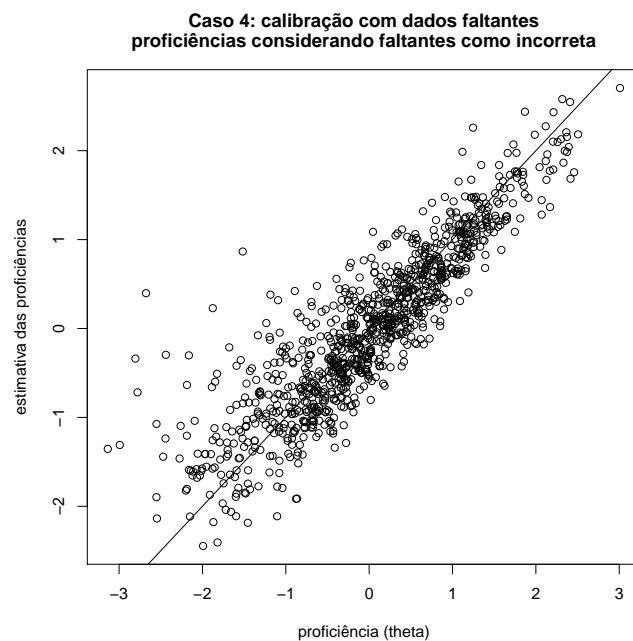


Figura 3.12: Caso 4: Simulação com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências. Diagrama de dispersão entre a proficiência gerada (valores reais dos parâmetros) e a proficiência estimada

Capítulo 4

Propostas para probabilidade de imputação

É correto supor que, para cada indivíduo, quanto maior for o grau de dificuldade do item, menor será a probabilidade de resposta positiva (resposta correta). Sendo assim, apresentamos dois procedimentos para imputação de dados faltantes, utilizando como base os valores estimados do parâmetro de dificuldade do item (b): imputação via regressão logística e imputação via regressão isotônica suavizada. Apesar de utilizar o parâmetro de dificuldade dos itens para cálculo da probabilidade de resposta positiva (resposta correta), nosso objetivo é melhorar as estimativas das proficiências.

Dos casos de estimação dos parâmetros apresentados no capítulo 3, o caso 4 é o mais utilizado em muitas avaliações que utilizam teoria de resposta ao item para determinar as proficiências dos alunos em pesquisas educacionais. Portanto, não apresentaremos simulações de imputação para os casos 2 e 3, mas apenas para o caso 4, que é o mais utilizado na prática.

No caso 4, do capítulo 3, foi realizada uma simulação com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências (com o parâmetro de acerto ao acaso fixo em 0.2). Ou seja, os dados faltantes foram tratados sem imputação, na calibração dos itens e com imputação de ‘zeros’, na estimação das proficiências. Neste capítulo, apresentamos dois procedimentos que, ao invés de tratar os dados faltantes como ignorados e depois como respostas incorretas, utilizaram imputação de dados, que fizeram

com que as estimativas dos parâmetros fossem feitas através de uma matriz completa de dados. As estimativas dos dados faltantes tratados sem imputação serviram para comparar com os dois procedimentos sugeridos de imputação de dados (porém, os 50 valores do parâmetro de acerto ao acaso c foram estimados, e não fixados em 0.2, como no capítulo anterior).

Neste capítulo, utilizando o caso 4 do capítulo 3, foram feitas comparações entre as estimativas da proficiência, da dificuldade do item e da discriminação do item, entre o modelo sem imputação com os modelos com imputação via regressão logística e via regressão isotônica suavizada.

Para cada um dos três procedimentos (sem imputação, com imputação via regressão logística e com imputação via regressão isotônica suavizada), os valores estimados dos parâmetros de proficiência, dificuldade do item e discriminação do item foram comparados com os valores dos parâmetros gerados aleatoriamente (dados reais) e com os valores dos parâmetros estimados através da matriz completa de ‘zeros’ e ‘uns’ já citada. Os valores estimados do parâmetro de acerto ao acaso, das estimativas com imputação, foram comparados com os valores das estimativas sem imputação e com os valores das estimativas feitas a partir da matriz completa de dados.

Também foram estimados o erro e o erro quadrático médio, como função da proficiência para os três métodos considerados (sem imputação, com imputação via regressão logística e com imputação via regressão isotônica suavizada).

Para as análises deste capítulo, utilizaremos os mesmos parâmetros reais gerados aleatoriamente e a mesma matriz de dados com mil respondentes gerada no capítulo 3 e, também, uma matriz gerada com 2000 respondentes, para analisar os vieses das estimativas dos parâmetros com aumento do tamanho da amostra. Porém, no capítulo 3 o valor do parâmetro de acertos ao acaso foi fixado em 0.2 e, neste capítulo, ele foi estimado, para cada um dos 50 itens.

A seguir, serão apresentados três procedimentos a serem comparados, utilizando o modelo do caso 4, do capítulo 3: estimação sem imputação (aqui, o parâmetro c foi estimado, portanto, os valores não serão os mesmos do capítulo anterior), estimação com imputação via regressão logística e via regressão isotônica suavizada.

4.1 Estimação sem imputação

Foram estimados 50 parâmetros de dificuldade do item , 50 parâmetros de discriminação, 50 parâmetros de acerto ao acaso (no capítulo 3, este parâmetro foi fixado em 0.2) e 1000 valores de proficiência, utilizando a matriz incompleta de dados (com cerca de 25% de dados faltantes), citada no capítulo 3. O modelo utilizado foi o mesmo do caso 4, do capítulo 3, um modelo de TRI de três parâmetros, com dados faltantes ignorados - no processo de calibração dos itens - e com dados faltantes como respostas incorretas - na estimação das proficiências. Ou seja, utilizando a matriz de dados incompleta, os parâmetros foram estimados sem utilizar métodos de imputação.

Após a estimação dos parâmetros, os valores encontrados foram comparados com os valores dos parâmetros gerados aleatoriamente (valores reais): 50 valores da distribuição uniforme $[0.2, 3]$, 50 valores distribuídos de forma equidistante no intervalo $[-2, 2]$ e 1000 valores da distribuição normal $N(0, 1)$, os mesmos valores do capítulo 3.

Além de comparar com os valores reais, os valores dos parâmetros estimados sem imputação também foram comparados com os valores estimados da matriz completa de dados, ou seja, a mesma matriz utilizada, porém, sem a retirada de cerca de 25% de dados.

As estimativas dos parâmetros sem imputação serão comparadas às duas propostas para imputação de dados faltantes a seguir, sendo que, após a imputação, teremos uma matriz de dados completa para estimação dos parâmetros, sem ter que codificar os dados faltantes como resposta incorreta na fase de estimação das proficiências.

4.1.1 Estimação sem imputação para os valores da proficiência

A figura 4.1 [a] apresenta um gráfico similar ao gráfico visto no capítulo 3, caso 4, figura 3.12. Porém, ao invés de tornar fixo o parâmetro de acerto ao acaso, foram estimados os 50 valores desse parâmetro, interferindo nas estimativas dos demais parâmetros. Na figura, tem-se que: no eixo horizontal estão 1000 valores de proficiências reais geradas aleatoriamente através de uma distribuição normal $N(0, 1)$ e no eixo vertical estão as estimativas das proficiência, considerando a mesma matriz incompleta do capítulo 3, com cerca de 25% de dados faltantes, onde foi realizada uma

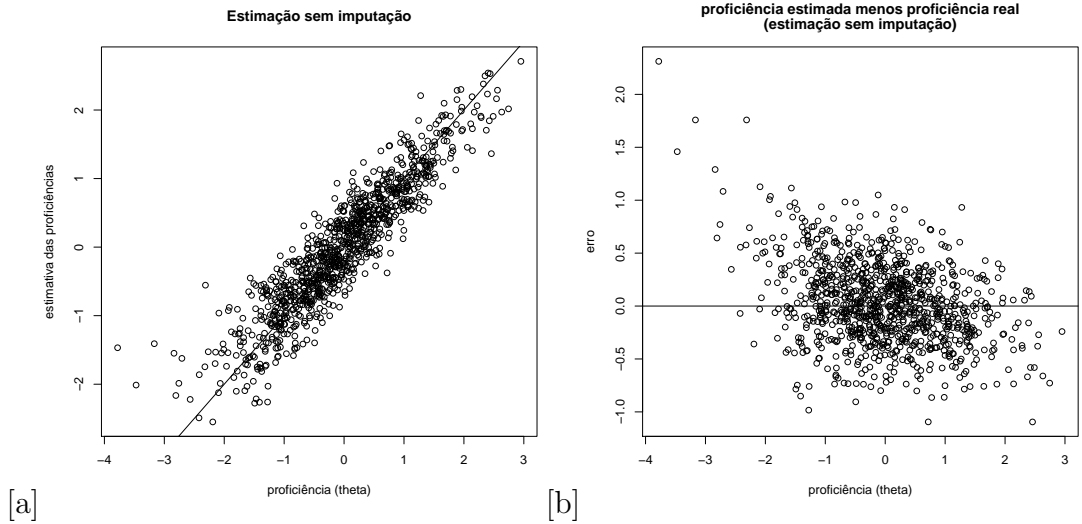


Figura 4.1: [a] Diagrama de dispersão entre a proficiência gerada (valores reais do parâmetro) e a proficiência estimada sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências). [b] diagrama de dispersão do erro entre os valores da proficiência gerada e a proficiência estimada.

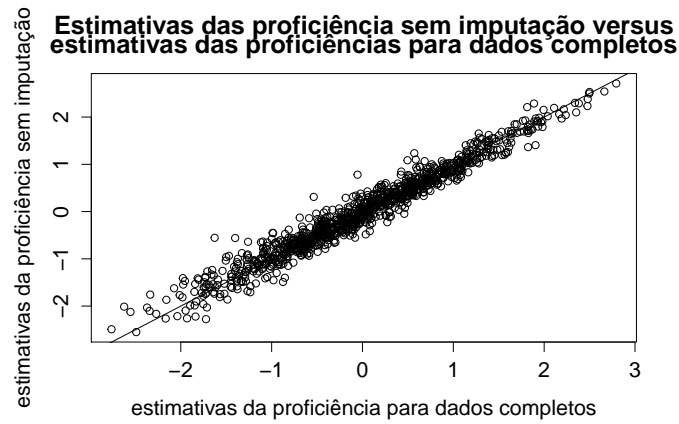


Figura 4.2: Diagrama de dispersão entre a proficiência estimada, através da matriz completa de dados, e a proficiência estimada sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).

simulação com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências. A figura

4.1 [b] apresenta um gráfico de dispersão do erro entre os dois eixos. Do mesmo modo apresentado no capítulo anterior, verifica-se a tendência a superestimação para os menores valores de θ .

A figura 4.2 mostra as estimativas da proficiência obtidas sem imputação versus as estimativas obtidas para dados completos, no caso de 1000 respondentes. Há alto grau de concordância entre as estimativas, especialmente para os casos de proficiência mais alta.

4.1.2 Estimação sem imputação para os valores do parâmetro de dificuldade dos itens

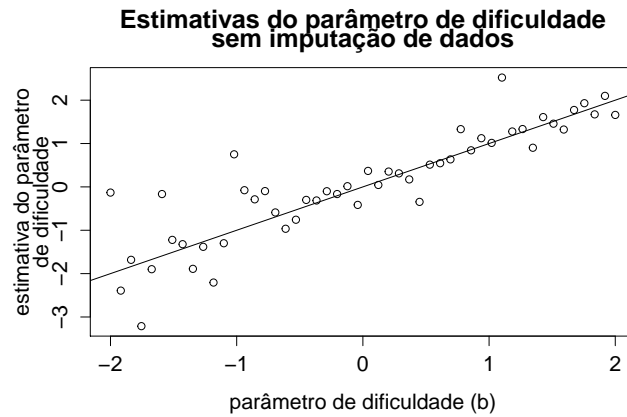


Figura 4.3: Diagrama de dispersão entre a dificuldade do item gerado (valores reais do parâmetro) e a dificuldade do item estimado sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).

A figura 4.3 mostra as estimativas do parâmetro de dificuldade do item obtidas sem imputação, no eixo vertical, para o caso de mil respondentes, e, no eixo horizontal, o valor real do parâmetro gerado aleatoriamente. O gráfico mostra alta concordância entre os valores para os itens de maior dificuldade.

A figura 4.4 mostra as estimativas do parâmetro de dificuldade sem imputação, com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências (com c estimado) versus os valores das estimativas através da matriz completa de dados, no caso de 1000 res-

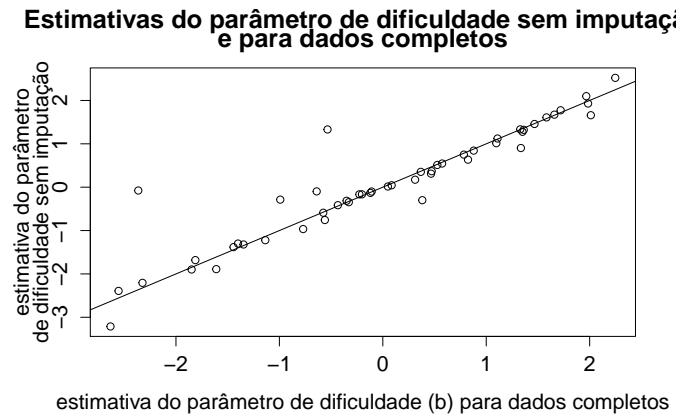


Figura 4.4: Diagrama de dispersão entre os valores das estimativas dos parâmetros de dificuldade do item estimado através da matriz completa de dados, e as estimativas do parâmetro de dificuldade do item sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).

pondentes. A discrepância nos valores estimados é pequena para quase todos os itens. Observa-se para valores reais (figura 4.3) maior discrepância do que as estimativas para dados completos.

4.1.3 Estimação sem imputação para os valores do parâmetro de discriminação dos itens

A figura 4.5 é análogo à figura 4.3, mas para o parâmetro de discriminação. Mostra pequena discrepância entre as estimativas do parâmetro de discriminação do item, obtidas sem imputação, em relação aos valores reais do parâmetro.

A figura 4.6 mostra as estimativas do parâmetro de discriminação, sem imputação de dados, versus os valores das estimativas para os dados completos, no caso de 1000 respondentes. A discrepância nos valores estimados é pequena para quase todos os itens.

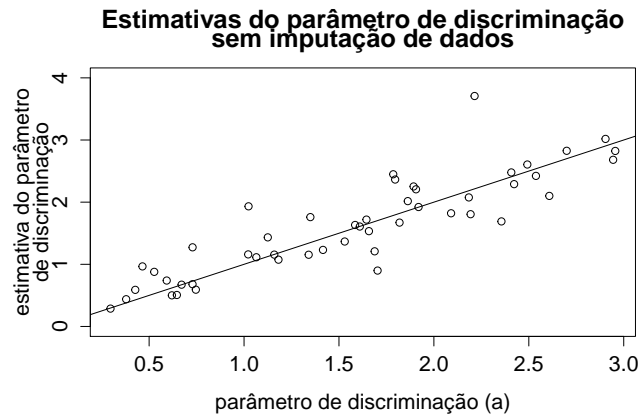


Figura 4.5: Diagrama de dispersão entre a discriminação do item gerado (valores reais do parâmetro) e a discriminação do item estimado sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).

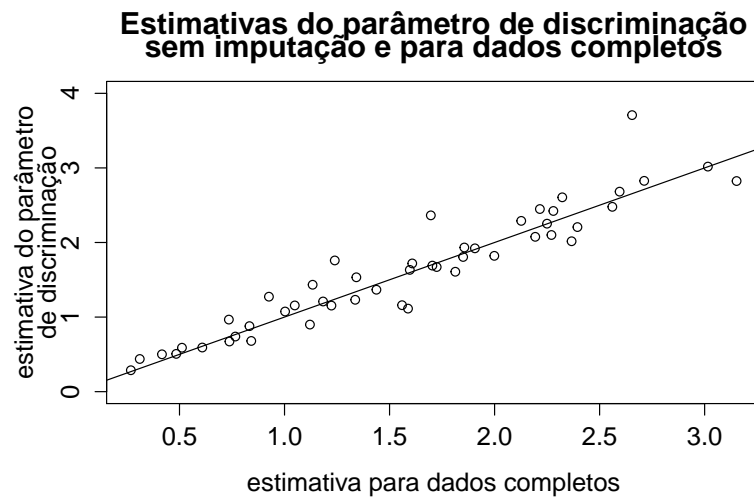


Figura 4.6: Diagrama de dispersão entre a discriminação do item estimado através da matriz completa de dados, e a discriminação do item estimado sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).

4.2 Exemplo de proposta para probabilidade de imputação utilizando regressão logística

Essa seção apresenta uma proposta de probabilidade de imputação utilizando a curva da regressão logística. As definições deste modelo estão no Apêndice.

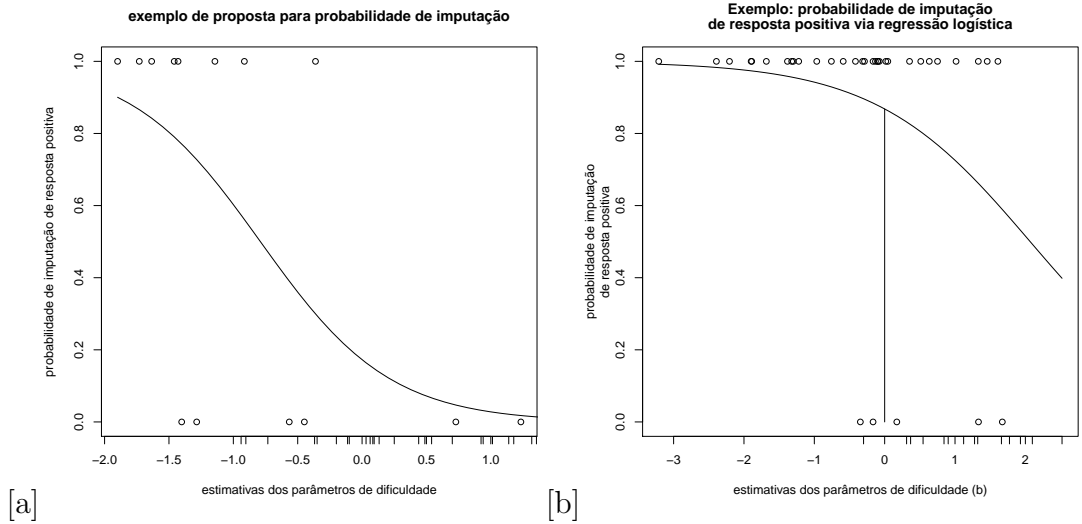


Figura 4.7: Exemplos de ajuste da curva logística para cálculo das probabilidades de imputação.

As figuras 4.7 [a] e 4.7 [b] mostram dois exemplos de ajuste da curva logística para cálculo das probabilidades de imputação para um indivíduo. Os pontos no gráfico correspondem às respostas corretas (1) e incorretas(0) do indivíduo. As marcas no eixo das abscissas correspondem aos valores estimados do parâmetro de dificuldade b dos itens não respondidos pelo indivíduo e para os quais a resposta será imputada com a probabilidade p_{ij} de imputação de resposta correta para o item não respondido k , dada pela curva logística obtida pelo ajuste do modelo:

$$\log \left(\frac{p_k}{1 - p_k} \right) = \beta_0 + \beta_1 \hat{b}_k \quad (4.1)$$

onde \hat{b}_k é o valor estimado do parâmetro de dificuldade b_k para o k -ésimo item para o qual o indivíduo forneceu uma resposta. O ajuste da curva logística foi elaborado utilizando as funções “tpm”, “eap”, “coef” e “factor.scores”, disponíveis na biblioteca “irtoys” do software R. Estas mesmas funções serão utilizadas nos estudos de simulação. A proposta de probabilidade de resposta correta a ser utilizada para imputação de dados faltantes foi obtida através da função “glm” do software R.

Utilizando-se o modelo logístico, a probabilidade de imputação de resposta correta é calculada pela fórmula:

$$p_{ij} = \frac{e^{d_j \hat{b}_i}}{1 + e^{d_j \hat{b}_i}} \quad (4.2)$$

onde:

p_{ij} = probabilidade de imputação de resposta positiva (ou correta) do j -ésimo indivíduo para acertar o i -ésimo item.

d_j = coeficiente de regressão logística para o j -ésimo indivíduo.

\hat{b}_i = estimativa do parâmetro de dificuldade do i -ésimo item.

Para gerar a resposta imputada U (0 ou 1, sendo 0 resposta errada e 1 resposta correta), utilizou-se a distribuição uniforme, onde:

Gerou-se o valor aleatório $y \sim \text{uniforme}(0, 1)$,

se $y \leq p_{ij} \Rightarrow U = 1$,

se $y > p_{ij} \Rightarrow U = 0$.

O processo de imputação foi repetido $m = 10$ vezes e adotamos como estimativa da proficiência θ_j do j -ésimo indivíduo a média das m estimativas obtidas após cada imputação.

Após a imputação dos valores, na matriz incompleta de dados do capítulo 3, encontrados via o modelo de regressão logística, foram estimados 50 valores do parâmetro de dificuldade do item, 50 valores do parâmetro de discriminação, 50 valores do parâmetro de acerto ao acaso e 1000 valores de proficiência. Esses valores foram comparados com os valores dos parâmetros gerados aleatoriamente: 50 valores da distribuição uniforme $[0.2, 3]$, 50 valores distribuídos de forma equidistante no intervalo $[-2, 2]$ e 1000 valores da distribuição normal $N(0, 1)$, os mesmos do capítulo 3. Através da matriz completa de ‘zeros’ e ‘uns’, citada no capítulo 3, caso 1, foram estimados 50 valores do parâmetro de dificuldade do item, 50 valores do parâmetro de discriminação, 50 valores do parâmetro de acerto ao acaso e 1000 valores de proficiência. Esses valores também foram comparados com as estimativas encontradas após imputação de dados via regressão logística.

Outra comparação realizada foi entre os valores encontrados nas estimativas com imputação via regressão logística e os valores encontrados sem imputação, onde foram estimados, através do modelo do caso 4, do capítulo 3, 50 valores do parâmetro de dificuldade do item, 50 valores do parâmetro de discriminação, 50 valores do

parâmetro de acerto ao acaso e 1000 valores de proficiência, utilizando a matriz incompleta de dados (com cerca de 25% de dados faltantes) citada anteriormente.

4.2.1 Estimação com imputação via regressão logística para os valores da proficiência

Após a imputação dos valores via regressão logística, utilizando-se a matriz completa dos dados, foram estimados os valores das proficiências dos indivíduos. A figura 4.8 [a] apresenta, no eixo horizontal, as proficiências geradas aleatoriamente através de uma distribuição normal $N(0, 1)$, a mesma utilizada no capítulo 3, e no eixo vertical apresenta as proficiências estimadas após imputação via regressão logística de uma matriz incompleta, a mesma simulada anteriormente com a retirada aleatória de cerca de 25% dos dados, no capítulo 3. A figura 4.8 [b] apresenta o gráfico dos erros entre os eixos verticais e horizontais. Com a imputação via regressão logística, observa-se a mesma tendenciosidade encontrada na simulação sem imputação (Figura 4.1).

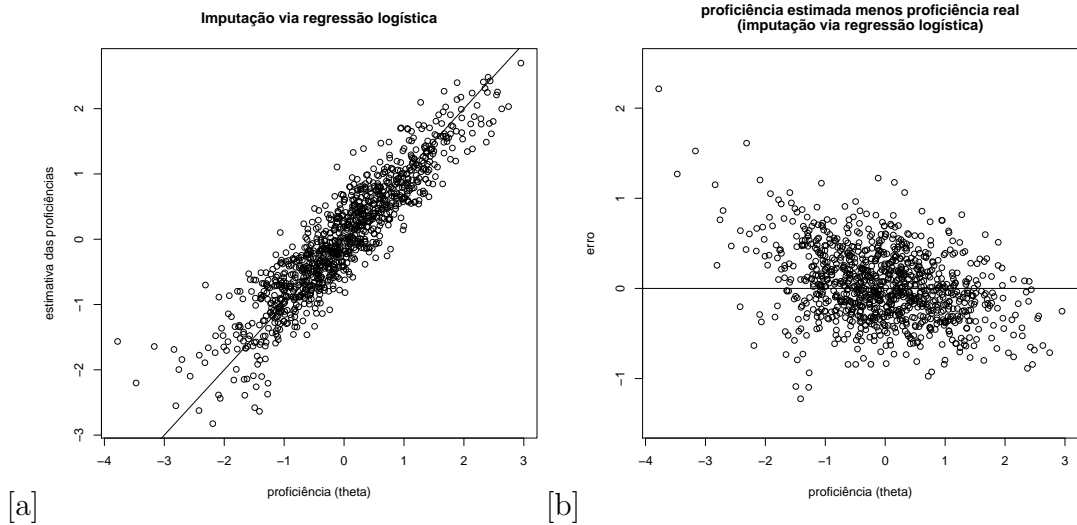


Figura 4.8: [a] Diagrama de dispersão entre os valores estimados da proficiência via regressão logística e a proficiência gerada (valores reais do parâmetro) [b] diagrama de dispersão do erro entre os valores estimados da proficiência via regressão logística e a proficiência gerada (valores reais do parâmetro).

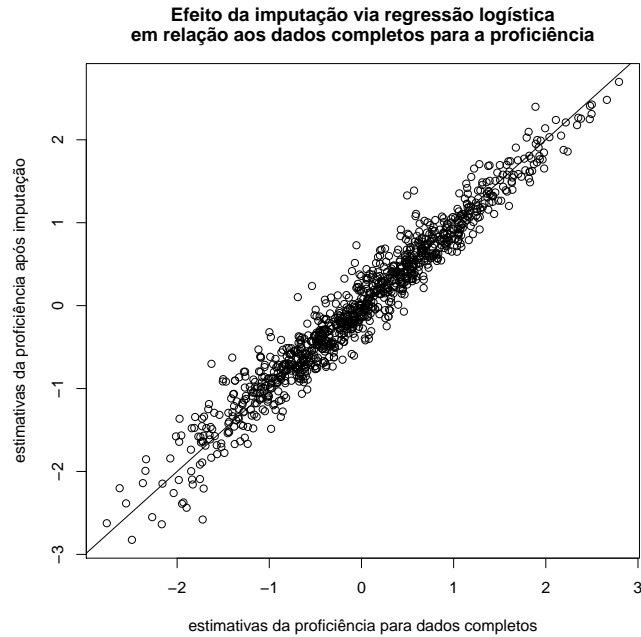


Figura 4.9: Diagrama de dispersão entre as estimativas da proficiência com imputação via regressão logística e os valores das estimativas do parâmetro da proficiência estimado através da matriz completa de dados.

A figura 4.9 mostra as estimativas da proficiência obtidas via imputação utilizando regressão logística e as estimativas obtidas para dados completos. Há alto grau de concordância entre as estimativas, especialmente para os casos de proficiência mais alta.

4.2.2 Estimação com imputação via regressão logística para os valores dos parâmetros de dificuldade dos itens

A figura 4.10 mostra as estimativas dos parâmetros de dificuldade obtidas com imputação via regressão logística (ordenadas) e os valores reais do parâmetro (abscissas), mostrando alta concordância entre os valores para os itens de maior dificuldade. Os valores mais baixos de dificuldade estão mais discrepantes.

O diagrama de dispersão (figura 4.11) apresenta as estimativas dos parâmetros sem imputação e com imputação via regressão logística. Apresenta alta concordância entre quase todos os valores.

A figura 4.12 é análoga à figura 4.4 e mostra maior discrepância entre os

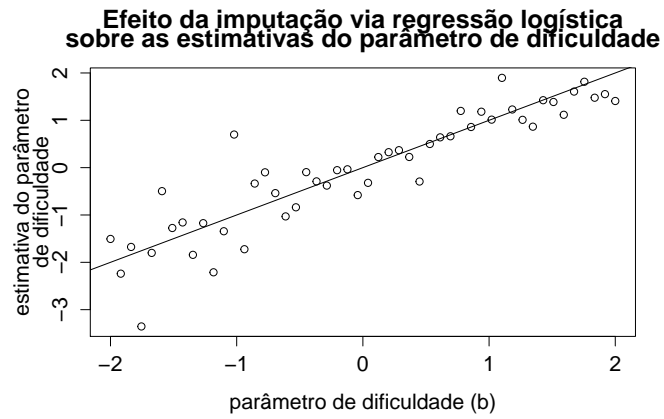


Figura 4.10: Diagrama de dispersão entre as estimativas do parâmetro de dificuldade com imputação via regressão logística e os valores das estimativas do parâmetro de dificuldade gerado (valores reais do parâmetro).

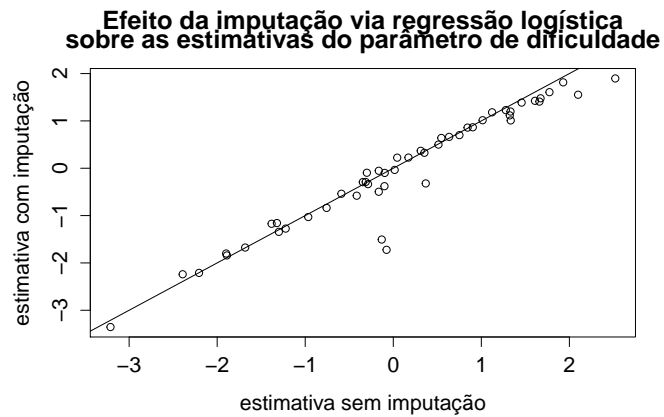


Figura 4.11: Diagrama de dispersão entre as estimativas do parâmetro de dificuldade com imputação via regressão logística e os valores das estimativas do parâmetro de dificuldade sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).

valores estimados do parâmetro de dificuldade com imputação via regressão logística em relação às estimativas para dados completos.

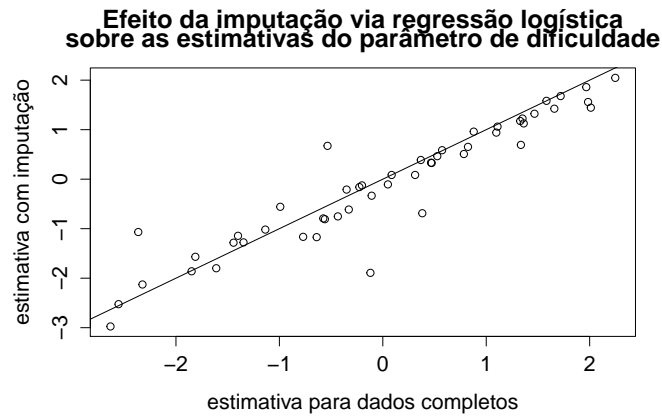


Figura 4.12: Diagrama de dispersão entre as estimativas do parâmetro de dificuldade com imputação via regressão logística e os valores das estimativas do parâmetro de dificuldade estimado através da matriz completa de dados.

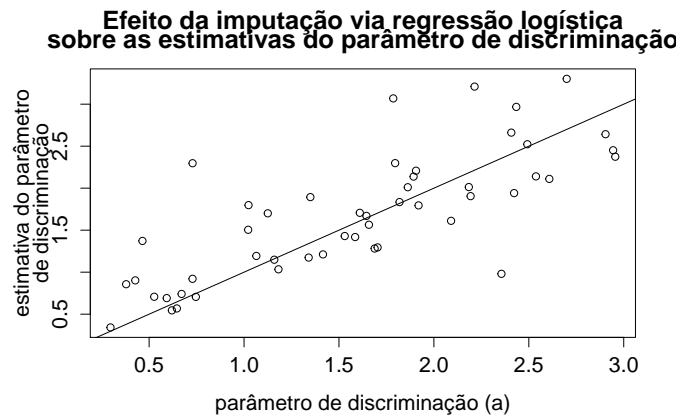


Figura 4.13: Diagrama de dispersão entre as estimativas do parâmetro de discriminação com imputação via regressão logística e o parâmetro de discriminação gerado (valores reais do parâmetro).

4.2.3 Estimação com imputação via regressão logística para os valores dos parâmetros de discriminação dos itens

A figura 4.13 mostra as estimativas do parâmetro de discriminação após imputação de dados via regressão logística versus os valores reais do parâmetro utilizados para geração dos dados, para o caso com 1000 respondentes.

A figura 4.14 mostra as estimativas dos parâmetros de discriminação obtidas com imputação via regressão logística (ordenadas) e as estimativas dos parâmetros de

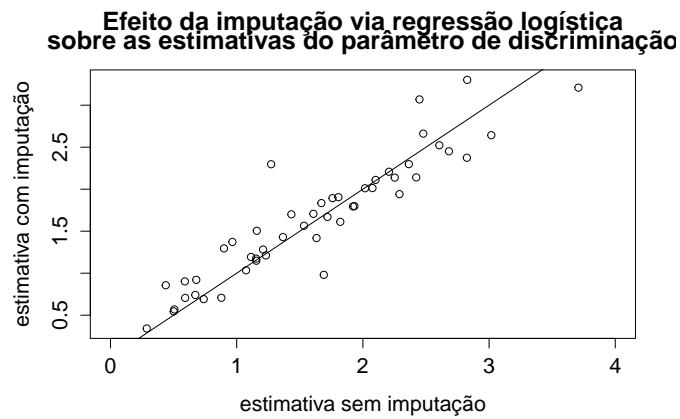


Figura 4.14: Diagrama de dispersão entre as estimativas do parâmetro de discriminação com imputação via regressão logística e os valores das estimativas do parâmetro de discriminação sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).

discriminação obtidas sem imputação (abscissas) para o caso com 1000 respondentes.

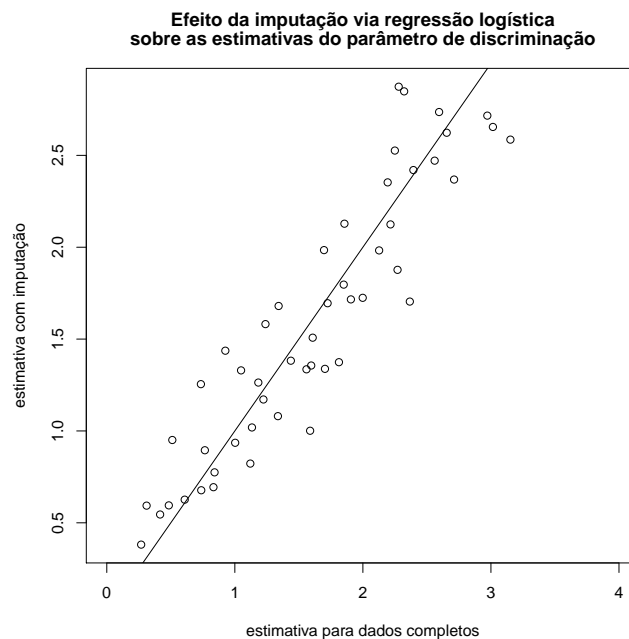


Figura 4.15: Diagrama de dispersão entre as estimativas do parâmetro de discriminação com imputação via regressão logística e os valores das estimativas do parâmetro de discriminação estimado através da matriz completa de dados.

A figura 4.15 mostra discrepância entre os valores estimados do parâmetro de discriminação com imputação via regressão logística em relação às estimativas para

dados completos.

As três estimativas (sem imputação, com imputação via regressão logística e para dados completos) apresentam valores próximos para a maioria dos itens.

4.2.4 Estimação com imputação via regressão logística para os valores dos parâmetros de acerto ao acaso

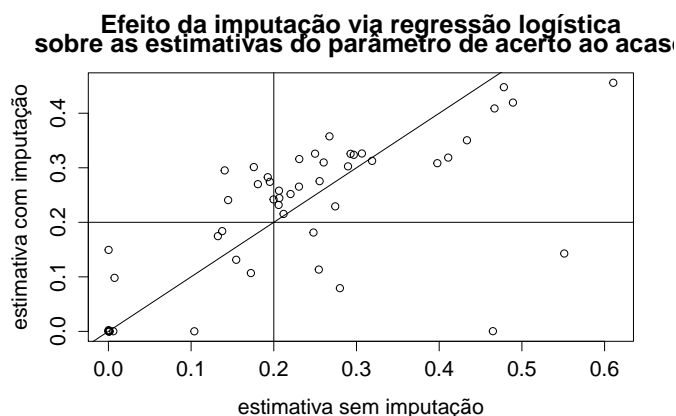


Figura 4.16: Diagrama de dispersão entre as estimativas do parâmetro de acerto ao acaso com imputação via regressão logística e os valores das estimativas do parâmetro de acerto ao acaso sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).

A figura 4.16 mostra as estimativas do parâmetro de acerto ao acaso obtidas após imputação de dados via regressão logística (ordenadas) e as estimativas obtidas sem imputação (abscissas) para cada um dos 50 itens, com 1000 respondentes. As estimativas obtidas sem imputação apresentam maior discrepância em relação ao valor real (0.2) do que as estimativas após imputação.

A figura 4.17 mostra as estimativas para o parâmetro de acerto ao acaso após imputação via regressão logística versus estimativas para dados completos no caso de 1000 respondentes.

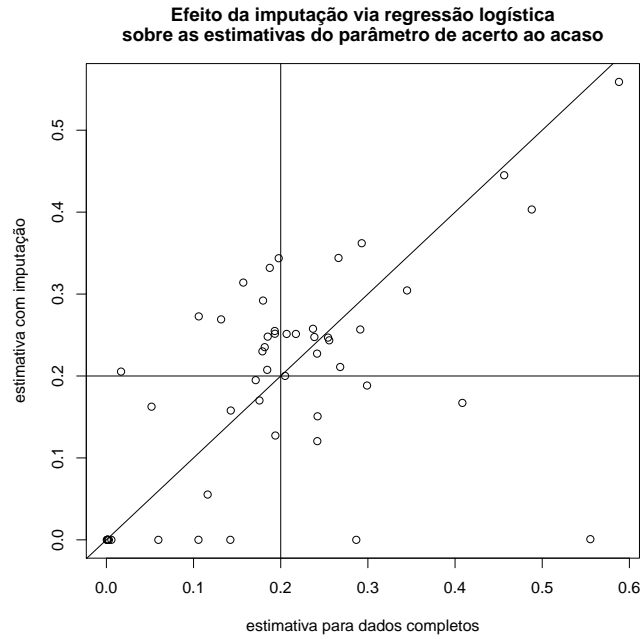


Figura 4.17: Diagrama de dispersão entre as estimativas do parâmetro de acerto ao acaso com imputação via regressão logística e os valores das estimativas do parâmetro de acerto ao acaso estimado através da matriz completa de dados.

4.3 Regressão isotônica

A regressão isotônica de uma função g observada nos pontos x_1, \dots, x_n é a função g^* que minimiza a soma de quadrados $\sum_{i=1}^n [g(x_i) - f(x_i)]^2$ no conjunto das funções f não decrescentes em x (assumimos $x_1 < x_2 < \dots < x_n$).

A regressão isotônica g^* no ponto x_i é dada pela derivada à esquerda do ponto i da função minorante convexa máxima do diagrama das somas acumuladas formado pelos pontos $(0, 0)$ e $p_i = (i, \sum_{j=1}^i g(x_j))$ $i = 1, \dots, n$ (figura 4.18[a]). Em nosso caso, para cada respondente, calculamos a regressão isotônica da função binária g indicadora de resposta positiva observada nos pontos dados pela estimativa do parâmetro de dificuldade em cada item. A probabilidade de resposta positiva deve decrescer quanto maior for o grau de dificuldade do item. Por este motivo, adotamos como probabilidade de imputação de resposta positiva a regressão isotônica obtida utilizando como função g as respostas aos itens em ordem inversa ao valor dos parâmetros de dificuldade dos itens.

Obtida a regressão isotônica, procedemos a sua suavização via kernel (núcleo

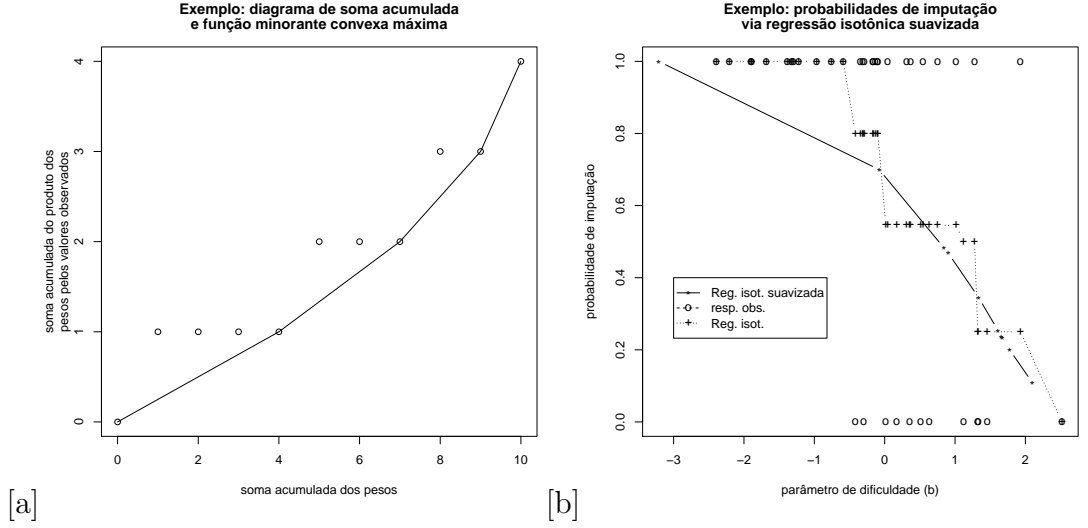


Figura 4.18: Exemplos de ajuste da curva isotônica para cálculo das probabilidades de imputação

estimadores) para obter probabilidades de imputação de resposta positiva para os itens não respondidos pelo indivíduo como função do grau de dificuldade dos itens não respondidos (figura 4.18[b]). A probabilidade de imputação de resposta positiva é dada pela expressão

$$p(\hat{b}_l) = \sum_{j=1}^{n_r} \Phi \left(\frac{\hat{b}_l - \hat{b}_j}{h} \right) \left[g^*(\hat{b}_j) - g^*(\hat{b}_{j-1}) \right], \quad l = 1, \dots, n_o,$$

onde $\Phi(\cdot)$ representa a função de distribuição de probabilidade acumulada da distribuição normal padrão, $g^*(\hat{b}_0) = 0$, n_r é o número de itens respondidos pelo indivíduo e n_o é o número de itens não respondidos pelo indivíduo. O valor do parâmetro de suavidade é dado por $h = 1.1/n^{0.2}$.

Assim como o modelo de regressão logística, após a imputação dos valores na matriz incompleta através do modelo de regressão isotônica suavizada, os parâmetros foram estimados e comparados com os valores dos parâmetros gerados (valores reais dos parâmetros) e com os valores estimados através da matriz completa de dados. Também foram comparados com os valores dos parâmetros da estimação sem imputação.

4.3.1 Estimação com imputação via regressão isotônica suavizada para os valores dos parâmetros da proficiência

Após a imputação dos valores via regressão isotônica foram estimados os valores das proficiências dos indivíduos utilizando-se do modelo de TRI proposto no caso 4 do capítulo 3. A figura 4.19[a] apresenta, no eixo vertical, as mesmas proficiências geradas aleatoriamente para uma matriz completa de dados do caso 1 e no eixo horizontal estão as estimativas das proficiências após imputação via regressão isotônica de uma matriz incompleta, a mesma simulada com a retirada aleatória de cerca de 25% dos dados no capítulo 3. A figura 4.19 [b] apresenta o gráfico dos erros entre os eixos verticais e horizontais. Com a imputação via regressão isotônica, observa-se a mesma tendenciosidade da simulação sem imputação e da simulação com imputação via regressão logística.

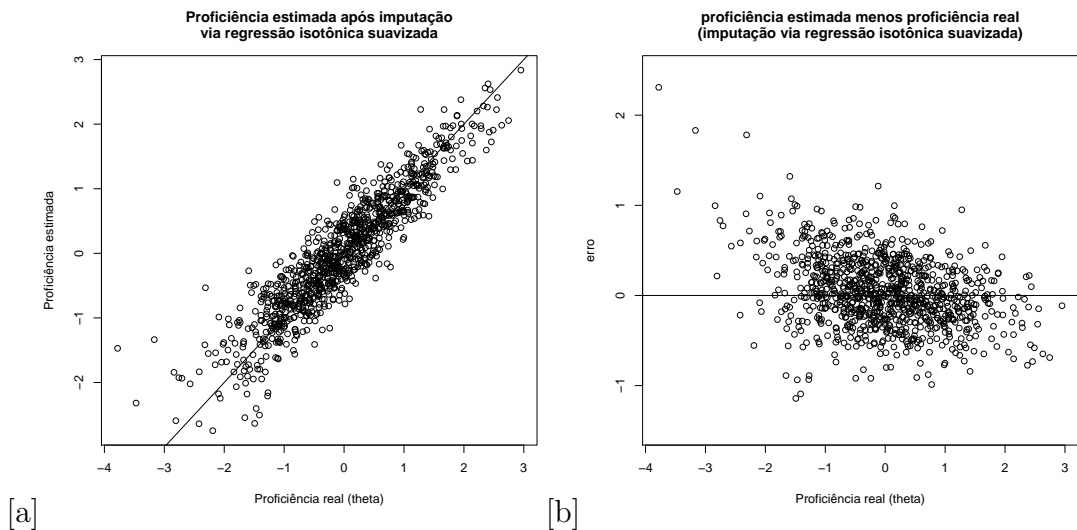


Figura 4.19: [a] Diagrama de dispersão entre os valores estimados da proficiência via regressão isotônica suavizada e a proficiência gerada (valores reais do parâmetro) [b] diagrama de dispersão do erro entre os valores os valores estimados da proficiência via regressão isotônica suavizada e a proficiência gerada (valores reais do parâmetro).

A figura 4.20 mostra as estimativas da proficiência obtidas via imputação utilizando regressão isotônica suavizada e as estimativas obtidas para dados completos. Há alto grau de concordância entre as estimativas, especialmente para os casos de

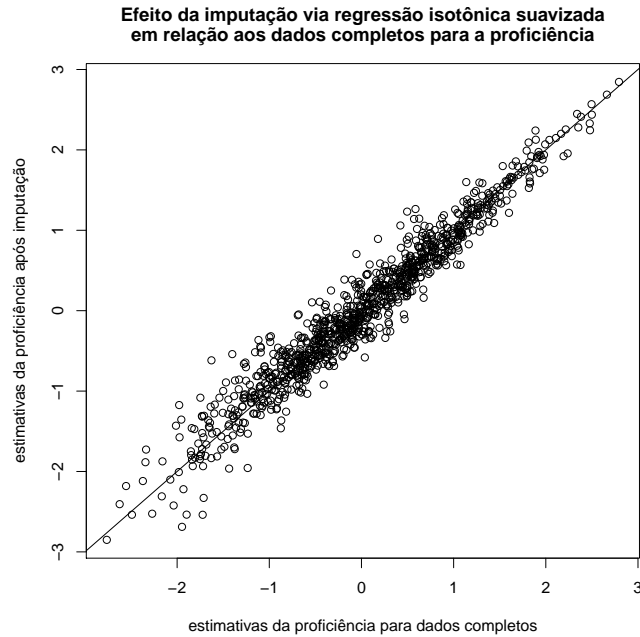


Figura 4.20: Diagrama de dispersão entre as estimativas da proficiência com imputação via regressão isotônica suavizada e os valores das estimativas de proficiência estimado através da matriz completa de dados.

proficiência mais alta. Nestes casos, a concordância entre as estimativas obtidas após imputação de dados via regressão isotônica suavizada e as estimativas para dados completos parece maior que o observado quando utilizamos a regressão logística para imputação.

4.3.2 Estimação com imputação via regressão isotônica suavizada para os valores dos parâmetros de dificuldade dos itens

Na figura 4.21, no eixo das ordenadas, estão as estimativas do parâmetro de dificuldade obtidas após imputação de dados via regressão isotônica suavizada, para o caso com 1000 respondentes. Mostra maior grau de concordância para os itens com maior dificuldade. Para valores menores, observa-se maior dispersão dos dados.

O diagrama de dispersão 4.22 apresenta as estimativas dos parâmetros sem imputação e com imputação via regressão isotônica suavizada.

A figura 4.23 mostra as estimativas do parâmetro de dificuldade após a im-

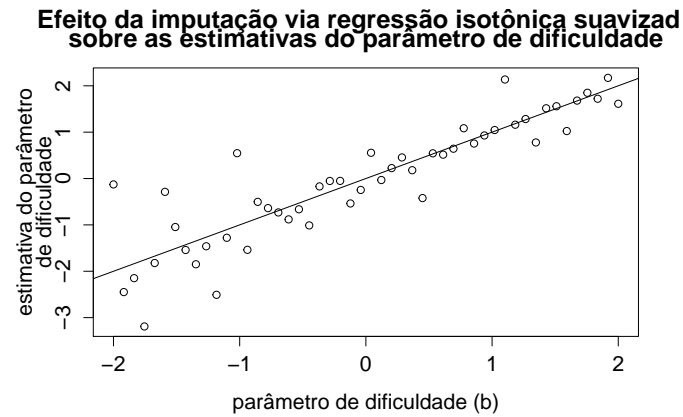


Figura 4.21: Diagrama de dispersão entre as estimativas do parâmetro de dificuldade com imputação via regressão isotônica suavizada e a dificuldade do item gerado (valores reais do parâmetro).

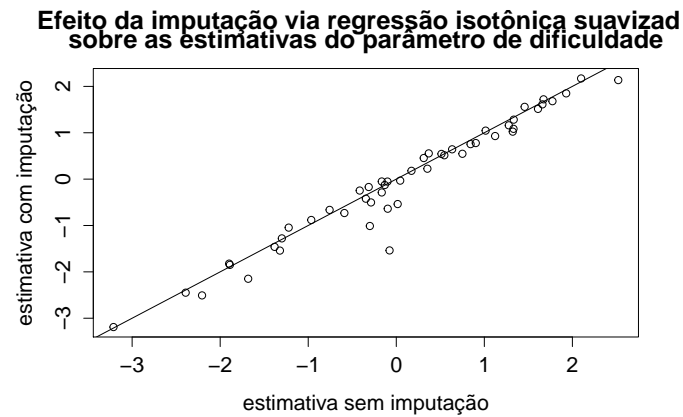


Figura 4.22: Diagrama de dispersão entre as estimativas do parâmetro de dificuldade com imputação via regressão isotônica suavizada e os valores das estimativas do parâmetro de dificuldade sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).

putação via regressão isotônica suavizada versus os valores das estimativas para os dados completos no caso de 1000 respondentes. A discrepância nos valores estimados é pequena para quase todos os itens.

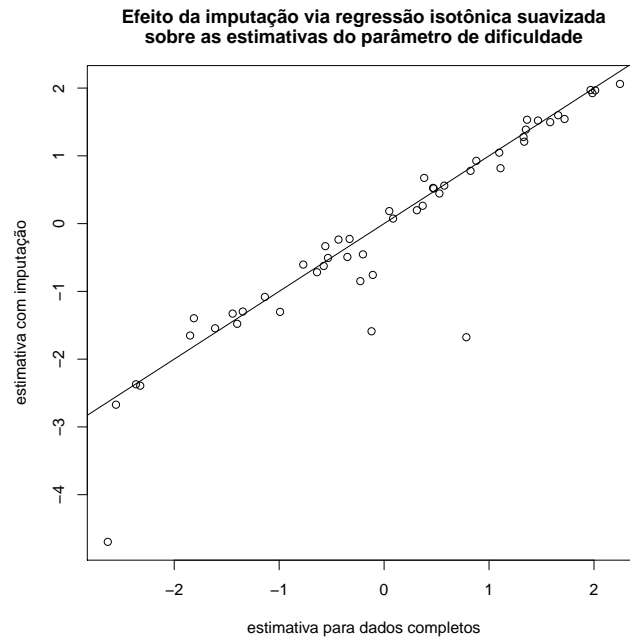


Figura 4.23: Diagrama de dispersão entre as estimativas do parâmetro de dificuldade com imputação via regressão isotônica suavizada e os valores das estimativas do parâmetro de dificuldade estimado através da matriz completa de dados.

4.3.3 Estimação com imputação via regressão isotônica suavizada para os valores dos parâmetros de discriminação dos itens.

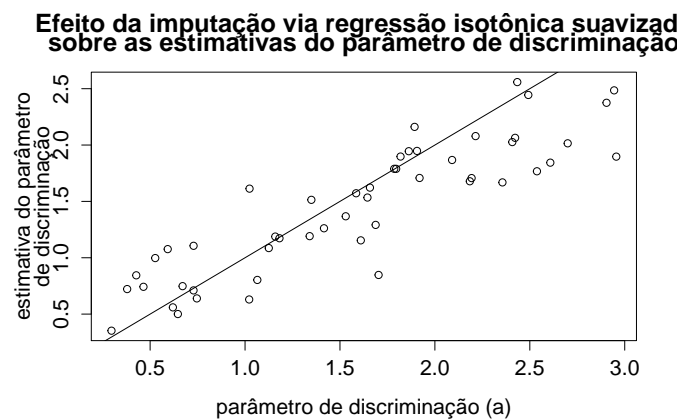


Figura 4.24: Diagrama de dispersão entre as estimativas do parâmetro de discriminação com imputação via regressão isotônica suavizada e discriminação do item gerado (valores reais do parâmetro).

A figura 4.24 mostra as estimativas do parâmetro de discriminação após a imputação via regressão isotônica suavizada versus o valor real do parâmetro utilizado na geração da matriz de respostas para o caso de 1000 respondentes. Mostra uma maior dispersão das estimativas dos valores mais altos, comparada com as estimativas obtidas sem imputação, com tendência de subestimação dos valores dos parâmetros para itens com maior grau de discriminação.

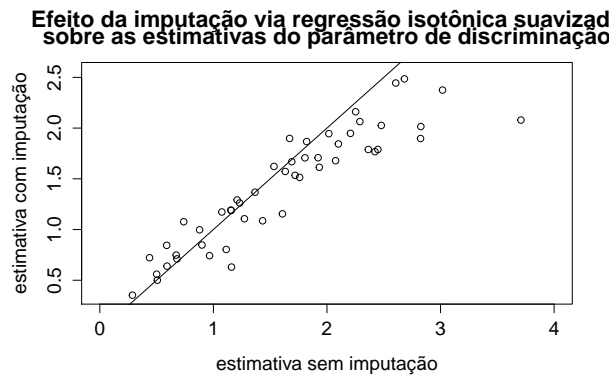


Figura 4.25: Diagrama de dispersão entre as estimativas do parâmetro de discriminação com imputação via regressão isotônica suavizada e os valores das estimativas do parâmetro de discriminação sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).

A figura 4.25 é análogo à figura 4.22, mas para o parâmetro de discriminação com 1000 respondentes. Mostra baixa discrepância para as estimativas mais baixas e maior para os itens com maiores estimativas para os valores do parâmetro, com as estimativas obtidas após imputação via regressão isotônica suavizada apresentando valores levemente menores do que os obtidos sem imputação.

A figura 4.26 mostra as estimativas do parâmetro de discriminação após imputação via regressão isotônica suavizada versus as estimativas para dados completos para o caso de 1000 respondentes. Há maior discrepância entre os valores para os itens com maior discriminação.

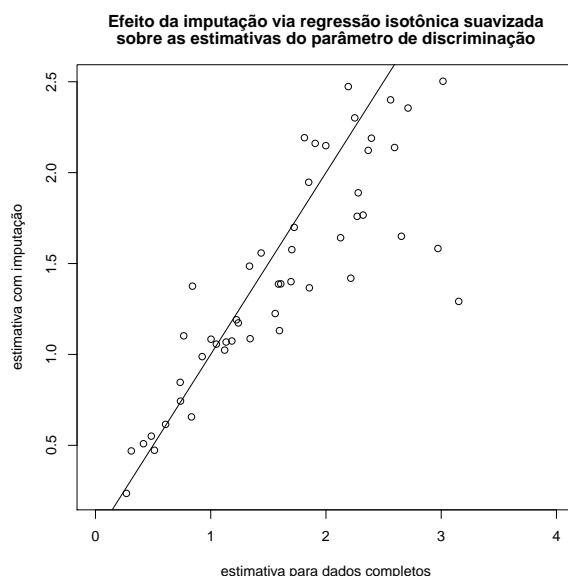


Figura 4.26: Diagrama de dispersão entre as estimativas do parâmetro de discriminação com imputação via regressão isotônica suavizada e os valores das estimativas do parâmetro de discriminação estimado através da matriz completa de dados.

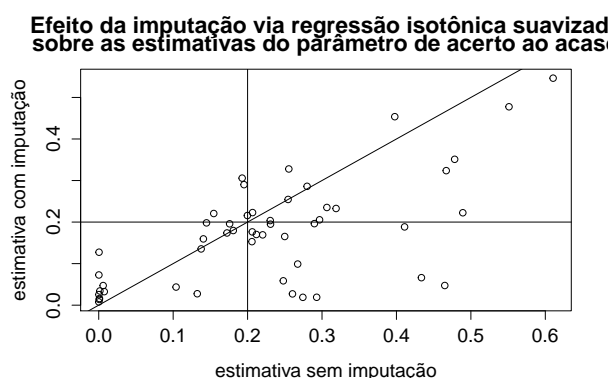


Figura 4.27: Diagrama de dispersão entre as estimativas do parâmetro de acerto ao acaso com imputação via regressão isotônica suavizada e os valores das estimativas do parâmetro de acerto ao acaso sem imputação (com dados faltantes ignorados no processo de calibração dos itens e com dados faltantes como respostas incorretas na estimação das proficiências).

4.3.4 Estimação com imputação via regressão isotônica suavizada para os valores dos parâmetros de acerto ao acaso.

A figura 4.27 mostra as estimativas do parâmetro de acerto ao acaso obtidas após imputação de dados via regressão isotônica suavizada (ordenadas) e as estimativas

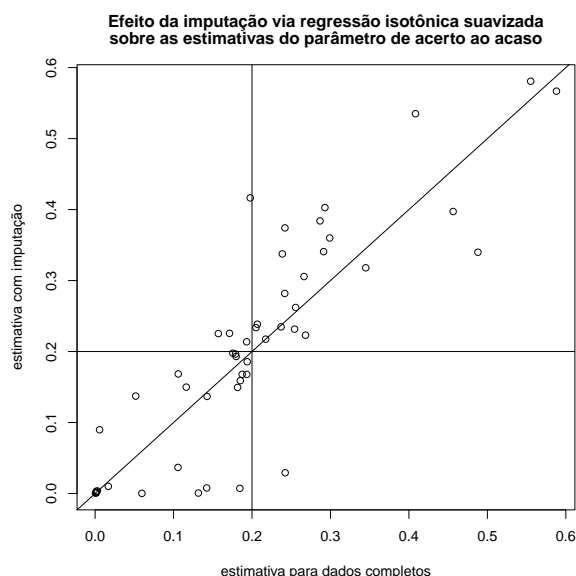


Figura 4.28: Diagrama de dispersão entre as estimativas do parâmetro de acerto ao acaso com imputação via regressão isotônica suavizada e os valores das estimativas do parâmetro de acerto ao acaso estimado através da matriz completa de dados.

obtidas sem imputação (abscissas) para cada um dos 50 itens para o caso de 1000 respondentes. Mostra uma tendência a obtermos valores menores do parâmetro após a imputação de dados via regressão isotônica suavizada. O valor 0.2 representa o valor real do parâmetro utilizado na geração dos dados simulados.

A figura 4.28 mostra as estimativas para o parâmetro de acerto ao acaso após imputação via regressão isotônica suavizada versus estimativas para dados completos no caso de 1000 respondentes. O padrão de discrepância em relação aos valores reais é similar nos dois casos.

4.4 Comparação entre os métodos de estimação das proficiências: sem imputação, com imputação via regressão logística e com imputação via regressão isotônica

:

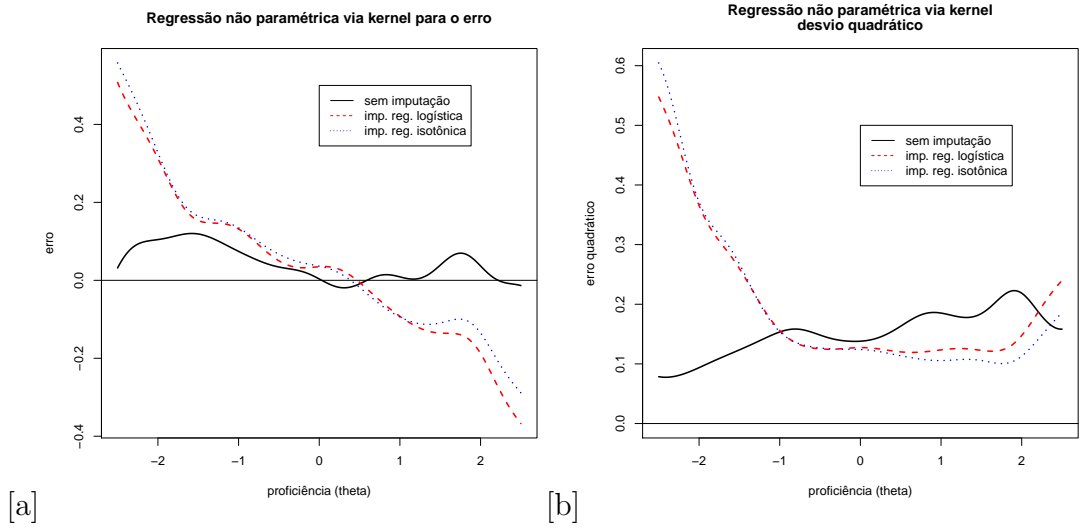


Figura 4.29: Regressão via kernel: [a] para o erro [b] desvio quadrático.

Para descrever o comportamento do erro e do erro quadrático na estimação da proficiência ao se utilizar os três métodos (sem imputação, com imputação via regressão logística, e com imputação via regressão isotônica suavizada via kernel), utilizamos a regressão não paramétrica via kernel de Nadaraya-Watson. A figura 4.29 [a] mostra, como função do possível valor da proficiência, a função

$$r(\theta) = \frac{\sum_{j=1}^n (\hat{\theta}_j - \theta_j) \phi\left(\frac{\theta - \theta_j}{h}\right)}{\sum_{j=1}^n \phi\left(\frac{\theta - \theta_j}{h}\right)}$$

onde $\phi(\cdot)$ representa a função densidade de probabilidade da distribuição normal padrão, n é o número de respondentes, e o valor do parâmetro de suavidade é dado por $h = 1.1/n^{0.2}$ (Ramsay, 1991).

Aqui, $\hat{\theta}_j$ representa o valor estimado da proficiência do j -ésimo respondente considerando cada um dos métodos utilizados na estimação (sem imputação, com im-

putação via regressão logística, e com imputação via regressão isotônica suavizada via kernel). Sendo assim, temos três curvas, que mostram que o erro das estimativas obtidas sem imputação de respostas faltantes mostra-se com melhor comportamento do que aquelas obtidas após imputação, com as estimativas via regressão isotônica suavizada apresentando melhor comportamento em relação à regressão logística para valores de θ maiores.

Com relação ao erro quadrático, a figura 4.29 [b] mostra a função

$$q(\theta) = \frac{\sum_{j=1}^n (\hat{\theta}_j - \theta_j)^2 \phi\left(\frac{\theta - \theta_j}{h}\right)}{\sum_{j=1}^n \phi\left(\frac{\theta - \theta_j}{h}\right)}$$

Podemos notar que as estimativas obtidas após imputação de respostas positivas apresenta melhores resultados, especialmente para a faixa de valores de proficiência acima de -1 , que tende a ser de maior interesse, sendo que a imputação via regressão isotônica suavizada apresenta curva mais próxima de zero do que a curva correspondente às estimativas obtidas com imputação via regressão logística.

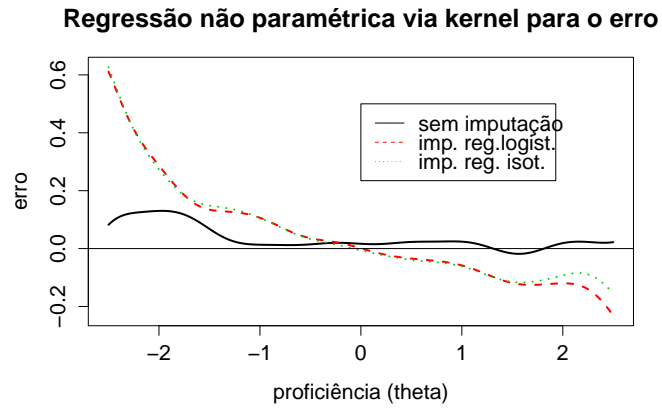


Figura 4.30: Regressão não paramétrica via kernel para o erro.

As figuras 4.30 e 4.31 são equivalentes às figuras 4.29 [a] e 4.29 [b], considerando 2000 respondentes. Mostram que o erro e o erro quadrático tendem a diminuir de magnitude para os três procedimentos com o aumento do tamanho do número de respondentes, como esperado. Nos dois casos, a imputação via regressão isotônica suavizada mostrou um padrão de valores mais baixos para o erro quadrático para os respondentes com maior proficiência.

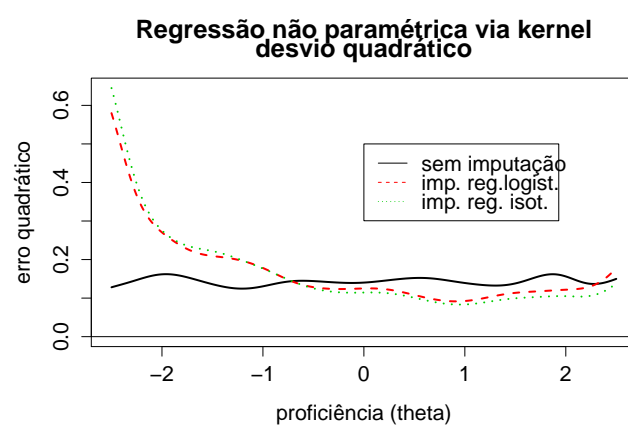


Figura 4.31: Regressão não paramétrica via kernel para o desvio quadrático.

Capítulo 5

Conclusões e Trabalhos Futuros

5.1 Conclusões

Os estudos de simulação mostraram que, para a estimação das proficiências, a utilização dos métodos de imputação de dados considerados apresentam erro de magnitude superior à estimação com dados faltantes, para amostra de tamanho 1000. No entanto, para amostras de 2000 respondentes, a magnitude do erro é pequena, sugerindo que os processos de imputação estudados podem ser adotados para casos com tamanho de amostras maiores. Com relação ao erro quadrático, as estimativas das proficiências apresentam valores menores do que para as estimativas com dados faltantes, especificamente para proficiências superiores a -1 (na escala da distribuição normal padrão), com o método de imputação via regressão isotônica suavizada apresentando os menores valores para o erro quadrático.

Quando não procedemos imputação de dados, os parâmetros de discriminação dos itens tendem a ser superestimados, enquanto os parâmetros de dificuldade são levemente subestimados. As proficiências tendem a ser superestimadas no caso das proficiências de menor valor.

Quando utilizamos imputação de dados via regressão logística e regressão isotônica suavizada via kernel, os parâmetros de discriminação e de dificuldade dos itens apresentam maior precisão em seus maiores valores. Além disso, as estimativas após imputação são muito similares às obtidas para dados completos.

5.2 Trabalhos Futuros

Outros estudos podem ser realizados utilizando-se amostras maiores. Sugere-se o estudar os procedimentos para o caso do modelo de 2 parâmetros.

Outra sugestão é a imputação via modelos politômicos, adontando a utilização do modelo de resposta gradual para dados politômicos onde uma categoria intermediária entre certo e errado é usada para não resposta.

Deve-se considerar, ainda, utilizar uma distribuição da proficiência com maior peso na cauda esquerda para obter melhores estimativas do parâmetro de acerto ao acaso na etapa de calibração dos itens, considerando-se posteriormente a população com distribuição normal para as proficiências, como é feito em estudos de avaliação educacional. Este procedimento deve melhorar as estimativas das proficiências.

- Andrade, D. F. de; Anjos, A. dos. (2012). Teoria da Resposta ao Item com uso do R. 20 SINAPE. Disponível em: <http://www.people.ufpr.br/aanjos/CE095/RTRIsinape.pdf>. Acesso em: 30/01/2014.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Baker, F. B. (1992), *Item Response Theory - Parameter Estimation Techniques.*, New York: Marcel Dekker, Inc.
- Baker, F.B. (2001). The basics of item response theory. Washington, DC: ERIC Clearinghouse on Assessment and Evaluation.
- Baker, R. (1987), *Classical Test Theory and Item Response Theory in Test Analysis.*, Special Report n.2: Language Testing Update. University of Edinburgh.
- Bejar, I. I. (1977). An application of the continuous response level model to personality measurement. *Applied Psychological Measurement*, 1, 509-521.
- Bethlehem, J. G (2002). Weighting adjustments for ignorable nonresponse. In Groves, R. M., Dillman, D. A., Eltinge, J. L., and Littler, R. J. A, editors, *Survey Nonresponse*, chapter 18, pages 275-287. John Wiley and Sons, New York.
- Birnbaum, A. (1968), Some latent trait models and their use in inferring an examinees ability. in f.m. lord & m. r. novick., in *Statistical Theories of Mental Test Scores*, Reding, MA: Addison-Wesley.
- Bock, R. D. & Lieberman, M. (1970). Fitting a response model for n dichotomousty scored items. *Psychometrika*, 35, 179-197.
- Bock, R. D. and Zimowski, M. F. (1997). Multiple Group IRT. In *Handbook of Modern Item Response Theory*. W.J. van der Linder e R.K. Hambleton Eds. New York: Springer-Verlag.
- Bock, R.D. e Aitkin M. (1981). Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. *Psychometrika*, 46, n. 4, 443-459.

- Book, R. D. (1972) Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- Christofferson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*. 40, 5-32.
- Cochran, W. G. (1977). *Sampling Techniques*. John Wiley and Sons, New York, 3rd edition.
- Colantonio, R. D. Pietro, A. Ocello, N. V. Verde. (2010). ABBA: Adaptive Bicluster-Based approach to impute missing values in binary matrices. *Proceedings of the 2010 ACM Symposium on Applied Computing*, pp. 1026-1033.
- Farhangfar, A.; Kurgan, L.; Pedrycz, W. (2007). A Novel Framework for Imputation of Missing Values in Databases. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 37(5): 692709.
- Finch, H. (2008). Estimation of item response theory parameters in the presence of missing data. *Journal of Educational Measurement*, 45, 225-245. doi:10.1111/j.1745-3984.2008.00062.x.
- Greenland, S. and Finkle, W. D. (1995). A critical look at methods for handling missing covariates in epidemiologic regression analyses. *American Journal of Epidemiology*, 142(12): 1255-1264.
- Hambleton, R. K. & Swaminathan, H. (1985), *Item Response Theory: Principles and Applications.*, Boston: Kluwer Academic Publishers.
- Hambleton, R. K. & Van Der Linden, W. J. (1996), *Modern Item Response Theory.*, Springer.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991), *Fundamentals of item response theory*, Newbury Park, CA: SAGE Publications.
- Horst, P. (1965). *Factor analysis of data matrices*. New York: Holt, Rinehart & Winston.

- Horton, N. J., & Kleinman, K. P. (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician*, 61, 7990. doi:10.1198/000313007X172556
- Huisman, M. (2000). Imputation of missing item responses: Some simple techniques. *Quality and Quantity*, 34, 331351. doi:10.1023/A:1004782230065
- king, G., Honaker, J., Joseph, A. Scheve, K. (2001). Analysing incomplete political science data: an alternative algorithm for multiple imputation. *American Political Science Review* 95, 49-69.
- Knol, M. J., Janssen, K. J. M., Donders, A. R. T., Donders, A. R. T., Egbert, A. C. G., Heerdink, E. R., Grobees, D. E., Moons, K. G. M., and Geerlings, M. I. (2010). Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example. *Journal of Clinical Epidemiology*, 63:728-736.
- Little, R. J. A. & Rubins, D.B (2002). *Statistical analysis with missing data*. 2 ed. New York: John Wiley & Sons.
- Lord, F. M. (1980), *Applications of item response theory to practical testing problems*, New Jersey: Lawrence Erlbaum Associates.
- Lord, F. M. (1968). An analysis of the Verbal Scholastic Aptitude Test using Birnbaum's threeparameter logistic model. *Educational and Psychological Measurement*, 28, 989-1020.
- Lord, F. M. (1953). An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability. *Psychometrika*, 18, 57-76.
- Lord, F.M. e Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. New York: Addison-Wesley.
- MacCorquidale K, Meehl PE (1948) On a distinction between hypothetical constructs and intervening variables. *Psychological Review* 55:95107.

- Marsh, H. W. (1998). Pairwise deletion for missing data in structural equation models: Nonpositive definite matrices, parameter estimates, goodness of fit, and adjusted sample size. *Structural Equation Modeling*, 5(1):22-36.
- McDonald, R.P. (1967). Nonlinear factor analysis. *Psychometric Monograph*, n. 15, 1-167.
- McDonald, R.P. (1985). *Factor analysis and related methods*. Hillsdale NJ: Erlbaum.
- McKinley, R. L. and Reckase, M. D. The use of the general rasch model with multidimensional item response data. Technical report, Iowa City, IA: American College Testing, 1980.
- McKnight PE, McKnight KM, Sidani S, Figueredo AJ (2007). *Missing data: A gentle introduction*. New York: Guilford.
- Miettien, O. S (1985). *Theoretical Epidemiology: Principles of Occurrence Research in Medicine*, John Wiley & Sons, New York.
- Molenberghs, G. and Kenward, M. G. (2007). *Missing Data in Clinical Studies*. John Wiley & Sons, Chichester, UK.
- Molenberghs, G., & Verbeke, G. (2004). An introduction to (generalized) (non)linear mixed models. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 111-153). New York: Springer
- Molenberghs, G., & Verbeke, G. (2005). *Models for discrete longitudinal data*. New York: Springer
- Muñiz, J. (1990). *Teoría de respuesta a los ítems: Un nuevo enfoque en la evolución psicológica y educativa*. Madrid: Ediciones Pirámide, S. A.
- Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551-560.

- Orchard, T. Woodbury, M. A. (1972). A Missing Information Principle: Theory and Applications, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, v. 1, pp. 697-715.
- Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen : Danish Institute for Educational Research.
- Rasch, G. (1962). On general laws and the meaning of measurement in psychology. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 4, 321-334.
- Reckase MD. (2009). Multidimensional item response theory. New York, NY: Springer.
- Rose, N., von Davier, M., & Wu, X. (2010). Modeling nonignorable missing data with item response theory (IRT) (ETS Research Report ETS-RR-10-11).
- Rubin, D. B (1996). Multiple imputation after 18+ years. Journal of the American Statistical Association, 91(43):473-489.
- Rubin, D. B. (1976). Inference and missing data. Biometrika, 63(3), 581-592.
- Rubin, D. B. (1987a). Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, New York.
- Samejima, F. (1974). Normal ogive model on the continuous response level in the multidimensional space. Psychometrika, 39, 111-121.
- Samejima, F. A. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometric Monograph, 17.
- Särndal, C. E., Swensson, B., and Wretman, J. (1992). Model Assisted Survey Sampling. Springer-Verlag, New York.
- Schafer, J. L. e Graham, J. W. (2002). Missing data: our view of the state of the art. Psychological Methods, 7, 147-177.
- Shafer, J. L. and Schenker, N. (2000). Inference with imputed conditional means. Journal of the American Statistical Association, 95(449):144-154.

- Sheng, X., & Carrière, K. C. (2005). Strategies for analyzing missing item response data with an application to lung cancer. *Biometrical Journal*, 47, 605-615.
- Sijtsma, K., & van der Ark, L. A. (2003). Investigation and treatment of missing item scores in test and questionnaire data. *Multivariate Behavioral Research*, 38, 505-528.
- Swaminathan, H. & Gifford, J. A. (1983), Estimation of Parameters in the Three-Parameter Latent Trait Model. In D. Weiss (Ed.), New York: Academic Press.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- Vach, W. (1994). *Logistic regression with missing values in the covariates*. New York: Springer-Verlag.
- Vach, W. and Blettner, M. (1991). Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. *American Journal of Epidemiology*, 134(8):895-907.
- Van Buuren, S. (2012). *Flexible Imputation of Missing Data*. Boca Raton, Florida: Chapman and Hall/CRC Press.
- van Ginkel, J. R., van der Ark, L. A., & Sijtsma, K. (2007). Multiple imputation of item scores in test and questionnaire data, and influence on psychometric results. *Multivariate Behavioral Research*, 42, 387-413. doi:10.1080/00273170701360803.
- Van Praag, B. M. S., Dijkstra, T. K., and Van Velzen, (1985). Least-squares theory based on general distributional assumptions with an application to the incomplete observations problem. *Psychometrika*, 98(1):86-90.
- Wauters, K., Desmet, P., & Van den Noortgate, W. (2010). Adaptive item-based learning environments based on the item response theory: Possibilities and challenges. *Journal of Computer Assisted Learning*, 26, 549-562. doi:10.1111/j.1365-2729.2010.00368.x.
- White, I. R. and Thompson, S. G. (2005). Adjusting for partially missing baseline measurements in randomized trial. *Statistics in Medicine*, 24(7):993-1007.

- Wingersky, M. S. & Lord, F. M. (1973). A computer program for estimating examinee ability and item characteristic curve parameters when there are omitted responses. Research Memorandum 73-2. Princeton, N. J.: Educational Testing Service.
- Wright, B.D. Stone, M. H. (2004). Making measures. Chicago: Phaneron Press.
- Wright. 9. D. (1977). Solving measurement problems with the Rasch model. Journal of Educational Measuremenst, 14,97-116.

Apêndice 1

5.2.1 Regressão Logística

No modelo logístico a variável resposta Y_i é binária. Uma variável binária assume dois valores, como por exemplo, $Y_i = 0$ e $Y_i = 1$, denominados "fracasso" e "sucesso", respectivamente. Neste caso, "sucesso" é o evento de interesse.

No modelo linear temos:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (5.1)$$

Assumindo que $E(\varepsilon_i) = 0$, obtemos que

$$E(Y_i) = \beta_0 + \beta_1 x_i. \quad (5.2)$$

A variável resposta Y tem distribuição Bernoulli (π), com probabilidade de sucesso $P(Y_i = 1) = \pi_i$ e de fracasso $P(Y_i = 0) = 1 - \pi_i$. Desta forma

$$E(Y_i) = \pi_i. \quad (5.3)$$

Igualando (2.2) e (2.3), temos

$$E(Y_i) = \pi_i = \beta_0 + \beta_1 x_i. \quad (5.4)$$

Essa igualdade viola as suposições do modelo linear. De fato,

i) Os erros não são normais, pois:

$y_i = 1 \Rightarrow \varepsilon_i = 1 - \beta_0 - \beta_1 x_i$, e $y_i = 0 \Rightarrow \varepsilon_i = 0 - \beta_0 - \beta_1 x_i$. Assim não faz sentido assumirmos a normalidade dos erros.

ii) Não homogeneidade da variância.

Temos que $\text{Var}(Y_i) = \pi_i(1 - \pi_i) = (\beta_0 + \beta_1 x_i)(1 - \beta_0 - \beta_1 x_i)$ então a variância de Y_i depende de x_i , e consequentemente, não é constante.

iii) Restrição para a resposta média $E(Y_i)$. Como a resposta média é obtida em probabilidades, temos que $0 \leq \beta_0 + \beta_1 x_i \leq 1$. Entretanto, esta restrição é inapropriada para resposta em um modelo linear, que assume valores no intervalo $(-\infty, \infty)$. Uma forma de resolver esse problema é utilizar o modelo logístico.

Muitas funções foram propostas para a análise de variáveis com respostas dicotômicas. Dentre elas a mais simples é a que dá origem ao modelo logístico. Do ponto de vista estatístico, este modelo é bastante flexível e de fácil interpretação.

Modelo Estatístico Um modelo de regressão logística simples é usado para o caso de regressão com uma variável explicativa.

Suponha uma amostra de n observações independentes da terna (x_i, m_i, y_i) , $i = 1, 2, \dots, n$, sendo que:

x_i é o valor da variável explicativa; m_i é a quantidade de itens verificados na amostra ; y_i número de ocorrência de um evento em m_i ensaios; e n é o tamanho da amostra. Com isso, assumimos que a variável resposta tem distribuição de probabilidade binomial ($Y_i \sim B(m_i, \pi_i)$), tal que

$$P[Y_i = y_i] = \binom{m_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{m_i - y_i}. \quad (5.5)$$

Para adequarmos a resposta média ao modelo linear, usamos a função de ligação

$$\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}, \quad i = 1, \dots, n, \quad (5.6)$$

que pode ser escrita como

$$\ln \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 x_i. \quad (5.7)$$

Utiliza-se o método da máxima verossimilhança para estimar os parâmetros (β_0, β_1) . De forma genérica, o método de máxima verossimilhança nos fornece valores para os parâmetros desconhecidos que maximizam a probabilidade de se obter determinado conjunto de dados.

Assumindo que $(x_0, m_0, y_0), \dots, (x_n, m_n, y_n)$ são independentes, a função de verossimilhança é da seguinte forma

$$\begin{aligned}
P[Y = y_1, \dots, y_n | \beta_0, \beta_1] &= \prod_{i=1}^n \binom{m_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{m_i - y_i} \\
&= \prod_{i=1}^n \binom{m_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{m_i} (1 - \pi_i)^{-y_i} \\
&= \prod_{i=1}^n \binom{m_i}{y_i} \frac{\pi_i^{y_i} (1 - \pi_i)^{m_i}}{(1 - \pi_i)^{y_i}} \\
&= \prod_{i=1}^n \binom{m_i}{y_i} \left(\frac{\pi_i}{1 - \pi_i} \right)^{y_i} (1 - \pi_i)^{m_i}
\end{aligned}$$

Ignorando o termo constante $\binom{m_i}{y_i}$, que não depende de β_0 e β_1 , e tomando o logaritmo (\ln) em ambos os lados da expressão anterior, temos

$$L(\beta_0, \beta_1 | (x_i; m_i; y_i)) = \ln \left(\frac{\pi_i}{1 - \pi_i} \right)^{y_i} + \ln (1 - \pi_i)^{m_i}$$

$$L(\beta_0, \beta_1 | (x_i; m_i; y_i)) = y_i \ln \left(\frac{\pi_i}{1 - \pi_i} \right) + m_i \ln(1 - \pi_i)$$

Detalhando $\ln \left(\frac{\pi_i}{1 - \pi_i} \right)$, e considerando que,

$$\pi_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}, \text{ temos}$$

$$\begin{aligned}
\ln \left(\frac{\pi_i}{1 - \pi_i} \right) &= \ln \left(\frac{\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}}{1 - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}} \right) \\
&= \ln \left(\frac{\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}}{\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}}} \right)
\end{aligned}$$

Assim a expressão, pode ser reescrita como:

$$\begin{aligned}
L(\beta_0, \beta_1 | (x_i; m_i; y_i)) &= \sum_{i=1}^n \left[y_i \ln \left(\frac{\pi_i}{1 - \pi_i} \right) + m_i \ln(1 - \pi_i) \right] \\
&= \sum_{i=1}^n \left[y_i (\beta_0 + \beta_1 x_i) + m_i \ln \left(1 - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \right]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \left[y_i (\beta_0 + \beta_1 x_i) + m_i \ln \left(\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \right] \\
&= \sum_{i=1}^n [y_i (\beta_0 + \beta_1 x_i) + m_i (\ln 1 - \ln(1 + e^{\beta_0 + \beta_1 x_i}))] \\
&= \sum_{i=1}^n y_i (\beta_0 + \beta_1 x_i) - \sum_{i=1}^n m_i \ln(1 + e^{\beta_0 + \beta_1 x_i})
\end{aligned}$$

Portanto,

$$L(\beta_0, \beta_1 | (x_i; m_i; y_i)) = \sum_{i=1}^n y_i (\beta_0 + \beta_1 x_i) - \sum_{i=1}^n m_i \ln(1 + e^{\beta_0 + \beta_1 x_i})$$

Para simplificar a notação faremos $L(\beta_0, \beta_1 | (x_i; m_i; y_i)) = L(\beta_0, \beta_1)$.

Estimação dos Parâmetros do modelo Para ajustar um modelo de regressão devemos estimar os parâmetros β_0 e β_1 do modelo. Os estimadores de máxima verossimilhança para os parâmetros β_0 e β_1 são os valores de $\hat{\beta}_0$ e $\hat{\beta}_1$ que maximizam o logaritmo da função de verossimilhança. A função de verossimilhança tem máximo, já que $0 < P[Y_i = y_i | x_i] < 1$, pois a função logaritmo é estritamente crescente.

Para maximizar a função de verossimilhança basta derivarmos em relação aos parâmetros do modelo, da seguinte forma

$$\begin{aligned}
\frac{\partial}{\partial \beta_0} L(\beta_0, \beta_1) &= \sum_{i=1}^n y_i - \sum_{i=1}^n m_i \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \\
\frac{\partial}{\partial \beta_1} L(\beta_0, \beta_1) &= \sum_{i=1}^n y_i x_i - \sum_{i=1}^n m_i x_i \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}
\end{aligned}$$

Igualando estas derivadas a zero e substituindo os parâmetros (β_0, β_1) pelos estimadores $(\hat{\beta}_0, \hat{\beta}_1)$, temos

$$\sum_{i=1}^n y_i - \sum_{i=1}^n m_i \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}} = 0$$

,

$$\sum_{i=1}^n y_i x_i - \sum_{i=1}^n m_i x_i \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}} = 0$$

. Porém, estas equações são não-lineares nos parâmetros e para resolvê-las é preciso recorrer a métodos numéricos iterativos, como Newton-Raphson (Gourieroux e Monfort, 1995). Este método é definido expandindo-se a função $U(\boldsymbol{\beta})$ em torno do ponto inicial $\boldsymbol{\beta}^{(0)}$, tal que

$$U(\boldsymbol{\beta}) \approx U(\boldsymbol{\beta}^{(0)}) + U'(\boldsymbol{\beta}^{(0)})(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}),$$

sendo que $U(\boldsymbol{\beta})$ é o vetor de derivadas de primeira ordem do logaritmo da função de verossimilhança em relação aos parâmetros do modelo e $U'(\boldsymbol{\beta})$ é a matriz de derivadas de ordem 2 do logaritmo da função de verossimilhança.

Se repetirmos o processo chegaremos ao processo iterativo

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + [-U'(\boldsymbol{\beta}^{(m)})]^{-1}U'(\boldsymbol{\beta}^{(m)}),$$

sendo que $m = 0, 1, \dots$

Como a matriz $-U'(\boldsymbol{\beta})$ pode não ser positiva definida, e portanto não inversível, ela é substituída pela matriz de informação de Fisher. Assim

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + [-I(\boldsymbol{\beta}^{(m)})^{-1}]U'(\boldsymbol{\beta}^{(m)}), \quad m = 0, 1, \dots$$

A matriz de informação de Fisher, para o modelo logístico com uma variável, tem a seguinte forma:

$$\begin{aligned} I(\hat{\boldsymbol{\beta}}) &= - \begin{bmatrix} \frac{\partial^2}{\partial \beta_0^2} \ln L(\beta_0, \beta_1) & \frac{\partial^2}{\partial \beta_0 \partial \beta_1} \ln L(\beta_0, \beta_1) \\ \frac{\partial^2}{\partial \beta_0 \partial \beta_1} \ln L(\beta_0, \beta_1) & \frac{\partial^2}{\partial \beta_1^2} \ln L(\beta_0, \beta_1) \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^n m_i \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} & \sum_{i=1}^n m_i x_i \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} \\ \sum_{i=1}^n m_i x_i \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} & \sum_{i=1}^n m_i x_i^2 \frac{e^{\beta_0 + \beta_1 x_i}}{(1 + e^{\beta_0 + \beta_1 x_i})^2} \end{bmatrix} \end{aligned}$$

Após obter as estimativas dos parâmetros do modelo é possível calcular as probabilidades estimadas

$$\hat{\pi}_i = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}$$

Interpretação dos parâmetros do modelo A interpretação dos parâmetros de um modelo de regressão logística é obtida comparando a probabilidade de sucesso

com a probabilidade de fracasso, usando a função odds ratio - OR (razão de chances). Essa função é obtida a partir da função odds.

$$g(x) = \frac{\pi(x)}{[1 - \pi(x)]} = \frac{\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}}{1 - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}} = \frac{\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}}{\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}}} = e^{\beta_0 + \beta_1 x_i}.$$

Assim, ao tomarmos dois valores distintos da variável explicativa, x_j e x_{j+1} , obtemos

$$OR = \frac{g(x_{j+1})}{g(x_j)} = \frac{e^{\beta_0 + \beta_1 x_{j+1}}}{e^{\beta_0 + \beta_1 x_j}}.$$

Temos ainda que:

$$\begin{aligned} \ln(OR) &= \ln \left[\frac{g(x_{j+1})}{g(x_j)} \right] = \ln [g(x_{j+1})] - \ln [g(x_j)] \\ &= \beta_0 + \beta_1 x_{j+1} - \beta_0 - \beta_1 x_j = \beta_1 (x_{j+1} - x_j). \end{aligned}$$

Fazendo $x_{j+1} - x_j = 1$ unidade, então

$$\ln(OR) = \ln(e^{\beta_1}) = \beta_1.$$

Assim, temos o quão provável o resultado ocorrerá entre os indivíduos x_{j+1} em relação aos indivíduos x_j , fazendo, portanto, algumas análises:

$$\beta_1 > 0 \Rightarrow OR > 1 \Rightarrow \pi(x_{j+1}) > \pi(x_j)$$

$$\beta_1 < 0 \Rightarrow OR < 1 \Rightarrow \pi(x_{j+1}) < \pi(x_j)$$

Estimativa dos desvios padrão No modelo de regressão logístico o desvio padrão dos estimadores é obtido a partir da matriz de informação de Fisher. Podemos ainda obter a matriz de informação de Fisher $I(\hat{\beta})$ para o modelo logístico a partir dos dados, da seguinte forma, $I(\hat{\beta}) = X' V X$, sendo que

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad V = \text{diag}[m_1 \hat{\pi}_1 (1 - \hat{\pi}_1), \dots, m_n \hat{\pi}_n (1 - \hat{\pi}_n)],$$

m_i é o número de repetições para cada elemento da amostra, $i = 1, \dots, n$.

As variâncias e covariâncias dos estimadores $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ são obtidos invertendo a matriz de informação de Fisher, isto é, calculando $\hat{\Sigma} = I^{-1}(\hat{\beta})$.

O j -ésimo elemento da diagonal principal da matriz $\hat{\Sigma}$ é a variância do estimador $\hat{\beta}_j$, denominada $\hat{\sigma}^2(\hat{\beta}_j)$. Os demais elementos da matriz $\hat{\Sigma}$ são as covariâncias entre $(\hat{\beta}_j; \hat{\beta}_u)$, $j \neq u$.

Desta forma o desvio padrão é definido como:

$$\widehat{DP}(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2(\hat{\beta}_j)}.$$

Inferência em um modelo logístico simples Após estimar os coeficientes, temos interesse em assegurar a significância das variáveis no modelo. Isto geralmente envolve formulação e teste de uma hipótese estatística para determinar se a variável independente no modelo é significativamente relacionada com a variável resposta. Para isso, temos os testes de hipóteses. Os testes de hipóteses mais utilizados são os testes da Razão da Verossimilhança, Wald e Escore.

Teste de Wald O teste de Wald é obtido por comparação entre a estimativa de máxima verossimilhança do parâmetro $(\hat{\beta}_1)$ e a estimativa de seu erro padrão. A razão resultante, sob a hipótese $H_0 : \beta_1 = 0$, tem distribuição normal padrão.

A estatística do teste Wald para a regressão logística é $W_j = \frac{\hat{\beta}_1}{\widehat{DP}(\hat{\beta}_1)}$.

O p-valor é definido como $P(|Z| > |W_j|)$, sendo que Z denota a variável aleatória da distribuição normal padrão.

Hauck e Donner examinaram o desempenho do teste de Wald e descobriram que ele se comporta de maneira estranha, em determinadas situações frequentemente não rejeitando a hipótese nula quando o coeficiente é significativo. Eles recomendam a utilização do teste da razão de verossimilhança para testar se realmente o coeficiente não é significativo quando o teste de Wald não rejeita a hipótese nula.

Teste da Razão de Verossimilhança Na regressão linear o interesse está no valor da SQR. Um valor alto da SQR sugere que a variável independente é importante, caso contrário, a variável independente não é útil na predição da variável resposta.

Na regressão logística a ideia é a mesma: comparar os valores observados da variável resposta com os valores preditos obtidos dos modelos com e sem a variável em questão. A comparação dos observados com os valores preditos é baseado no log da verossimilhança. Para entender melhor essa comparação, é útil pensar em um valor observado da variável resposta também como sendo um valor predito resultante de um modelo saturado. Um modelo saturado é aquele que contém tantos parâmetros quanto observações.

A comparação dos observados com os valores preditos usando a função de verossimilhança é baseada na seguinte expressão:

$$D = -2\ln \left[\frac{(\text{verossimilhança do modelo ajustado})}{(\text{verossimilhança do modelo saturado})} \right].$$

Com o propósito de assegurar a significância de uma variável independente, comparamos o valor da D com e sem a variável na equação. A mudança em D devido a inclusão da variável no modelo é obtida da seguinte maneira:

$$G = D(\text{modelo sem a variável}) - D(\text{modelo com a variável}).$$

Podemos então escrever a estatística G como:

$$G = -2\ln \left[\frac{(\text{verossimilhança sem a variável})}{(\text{verossimilhança com a variável})} \right].$$

ou ainda:

$$G = -2\ln(L_s) + 2\ln(L_c),$$

em que L_s é a verossimilhança do modelo sem a covariável e L_c é a verossimilhança do modelo com a covariável.

Queremos testar:

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

Sob a hipótese nula, a estatística G tem distribuição chi-quadrado com 1 grau de liberdade.

Teste Score A estatística teste para o Teste Score é:

$$ST = \frac{\sum_{i=1}^n x_i(y_i - \bar{y})}{[\bar{y}(1 - \bar{y}) \sum_{i=1}^n (x_i - \bar{x})^2]^{1/2}},$$

em que $\bar{y} = \hat{\pi}$ (proporção de sucessos na amostra).

No Teste Score também temos o interesse em testar:

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

O p-valor é definido como $P(|Z| > |ST|)$, sendo que Z denota uma variável aleatória com distribuição normal padrão.

Intervalo de Confiança para os parâmetros

A base da construção do intervalo de confiança para os parâmetros é a mesma teoria estatística que usamos para os testes de significância do modelo. Em particular, um intervalo de confiança para a inclinação e intercepto são baseados em seus respectivos testes de Wald. O intervalo de confiança de $100(1 - \alpha)\%$ para o parâmetro β_1 é:

$$IC(\beta_1, 1 - \alpha) = [\hat{\beta}_1 - z_{1-\alpha/2} DP(\hat{\beta}_1); \quad \hat{\beta}_1 + z_{1-\alpha/2} DP(\hat{\beta}_1)].$$

E para o intercepto:

$$IC(\beta_0, 1 - \alpha) = [\hat{\beta}_0 - z_{1-\alpha/2} DP(\hat{\beta}_0); \quad \hat{\beta}_0 + z_{1-\alpha/2} DP(\hat{\beta}_0)],$$

em que $z_{1-\alpha/2}$ é o ponto da normal padrão correspondente a $100(1 - \alpha/2)\%$.

Intervalo de Confiança para Logito

A logito é a parte linear do modelo de regressão logística. O estimador para logito é:

$$\hat{g}(x) = \hat{\beta}_0 + \hat{\beta}_1 x.$$

O estimador da variância do estimador da logito requer a obtenção da variância da soma. No caso é:

$$\widehat{Var}[\hat{g}(x)] = \widehat{Var}(\hat{\beta}_0) + x^2 \widehat{Var}(\hat{\beta}_1) + 2x \widehat{Cov}(\hat{\beta}_0, \hat{\beta}_1).$$

O intervalo de confiança para a logito é:

$$IC(g(x), 1 - \alpha) = [\hat{g}(x) - z_{1-\alpha/2} DP(\hat{g}(x)); \quad \hat{g}(x) + z_{1-\alpha/2} DP(\hat{g}(x))],$$

em que $DP(\hat{g}(x))$ é a raiz quadrada e $z_{1-\alpha/2}$ é o ponto da normal padrão.

Intervalo de Confiança para os valores ajustados

O estimador do logito e seu intervalo de confiança fornece o estimador dos valores ajustados. O intervalo de confiança dos valores ajustados é dado por:

$$IC(\pi, 1 - \alpha) = \left[\frac{e^{\hat{g}(x) - z_{1-\alpha/2} DP(\hat{g}(x))}}{1 + e^{\hat{g}(x) - z_{1-\alpha/2} DP(\hat{g}(x))}}; \quad \frac{e^{\hat{g}(x) + z_{1-\alpha/2} DP(\hat{g}(x))}}{1 + e^{\hat{g}(x) + z_{1-\alpha/2} DP(\hat{g}(x))}} \right].$$

4.1.4.4 Intervalo de Confiança para a Odds Ratio

Sejam os limites do intervalo de confiança para β_1 :

$$\beta_I = \hat{\beta}_1 - z_{1-\alpha/2} DP(\hat{\beta}_1) \text{ e } \beta_S = \hat{\beta}_1 + z_{1-\alpha/2} DP(\hat{\beta}_1).$$

O intervalo de confiança para a Odds Ratio é:

$$IC(Odds \text{ Ratio}, 1 - \alpha) = [e^{\beta_I}; \quad e^{\beta_S}].$$

Regressão Logística Múltipla Apresentamos na Seção anterior o modelo de regressão logística considerando apenas uma variável explicativa. Assim como no modelo de regressão linear, podemos ajustar um modelo para a variável resposta levando em conta mais de uma variável explicativa (covariável), o que chamamos de Modelo de Regressão Logística Múltipla.

Assim, um modelo de regressão logística múltipla é usado para o caso de regressão com mais de uma variável explicativa.

Considere um conjunto de p variáveis independentes denotadas como um vetor $X = (X_1, X_2, X_3, \dots, X_p)$. Neste caso a função de ligação é

$$g(X) = \ln \left(\frac{\pi(X)}{1 - \pi(X)} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

e

$$E[Y] = \pi(X) = \frac{e^{g(X)}}{1 + e^{g(X)}}.$$

Estimativas dos parâmetros do modelo Para obter as estimativas para vetor $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ dos parâmetros do modelo e a matriz de covariâncias será utilizado o método de máxima verossimilhança, da seguinte forma:

$$\mathbf{L}(\beta_0, \beta_1, \dots, \beta_p | (x_i; m_i; y_i)) = \sum_{i=1}^n [y_i g(X) - m_i \ln(1 + e^{g(X)})]$$

Derivando em relação aos parâmetros, temos:

$$\frac{\partial}{\partial \beta_0} \mathbf{L}(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^n y_i - \sum_{i=1}^n m_i \frac{e^{g(X)}}{1 + e^{g(X)}}.$$

$$\frac{\partial}{\partial \beta_j} \mathbf{L}(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^n y_i x_j - \sum_{i=1}^n m_i x_j \frac{e^{g(X)}}{1 + e^{g(X)}}.$$

Igualando a zero e substituindo os parâmetros pelos estimadores, tem-se as seguintes equações:

$$\sum_{i=1}^n y_i (1 + e^{g(X)}) - \sum_{i=1}^n m_i e^{g(X)} = 0$$

$$\sum_{i=1}^n y_i x_i (1 + e^{g(X)}) - \sum_{i=1}^n m_i x_i e^{g(X)} = 0$$

A solução dessas equações fornece as estimativas dos parâmetros do modelo, utilizando processo iterativo, análogo à estimação dos parâmetros do modelo de regressão logística simples.

Após obter as estimativas dos parâmetros do modelo, podemos calcular as probabilidades ajustadas:

$$\hat{\pi}_i = \frac{e^{\hat{g}(X_i)}}{1 + e^{\hat{g}(X_i)}},$$

em que

$$\hat{g}(X_i) = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_p X_{pi}.$$

Estimativa do desvio padrão O método de estimar as variâncias e covariâncias dos coeficientes estimados segue a teoria da estimação de máxima verossimilhança. Essa teoria assegura que os estimadores são obtidos da matriz de segundas derivadas parciais da função log verossimilhança. Essas derivadas parciais têm a seguinte forma geral:

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = - \sum_{i=1}^n m_i x_{ij}^2 \pi_i (1 - \pi_i).$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_l} = - \sum_{i=1}^n m_i x_{ij} x_{il} \pi_i (1 - \pi_i)$$

para $j, l=0, 1, \dots, p$ em que π_i simplifica $\pi(x_i)$. Seja a matriz $(p+1) \times (p+1)$ que contém os termos negativos denotada por $I(\beta)$ que é matriz de informação de Fisher. As variâncias e covariâncias dos coeficientes estimados são obtidos da inversa da matriz: $Var(\beta) = I^{-1}(\beta)$. A variância de β_j , $Var(\beta_j)$, é o j^{th} elemento da diagonal da matriz e a $Cov(\beta_j, \beta_l)$ é obtida através do elemento da matriz referente a linha de β_j e coluna de β_l ou vice-versa, já que $Cov(\beta_j, \beta_l) = Cov(\beta_l, \beta_j)$. Os estimadores das variâncias e covariâncias, $\hat{Var}(\hat{\beta})$, são obtidos de $Var(\beta)$ em $\hat{\beta}$.

Ainda, a matriz de informação de Fisher estimada pode ser obtida por: $\hat{I}(\hat{\beta}) = X'VX$.

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ & & \ddots & \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}_{n \times (p+1)}$$

e

$$V = \begin{bmatrix} m_1 \hat{\pi}_1 (1 - \hat{\pi}_1) & 0 & \dots & 0 \\ 0 & m_2 \hat{\pi}_2 (1 - \hat{\pi}_2) & \dots & 0 \\ & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & m_n \hat{\pi}_n (1 - \hat{\pi}_n) \end{bmatrix}_{n \times n}$$

Dessa forma, o desvio padrão do coeficiente β_j é:

$$\hat{DP}(\hat{\beta}_j) = \sqrt{\hat{Var}(\hat{\beta}_j)}.$$

Inferência no modelo de regressão logística múltipla Assim como na Inferência em um modelo de regressão logística simples, podemos testar a significância dos parâmetros pelo teste de Wald, score e Razão de Verossimilhança (TRV).

Teste da Razão de Verossimilhança

O teste da razão de verossimilhança para a significância dos p coeficientes das variáveis independentes do modelo é realizado da mesma maneira que no modelo de regressão logística simples. A estatística teste G é dada por:

$$G = -2\ln \left[\frac{(\text{verossimilhança sem a variável})}{(\text{verossimilhança com a variável})} \right].$$

ou ainda:

$$G = -2\ln(L_s) + 2\ln(L_c),$$

em que L_s é a verossimilhança do modelo sem a covariável e L_c é a verossimilhança do modelo com a covariável.

No caso da regressão múltipla, temos o interesse em saber se pelo menos uma variável é significativa para o modelo. Sob a hipótese nula, os p coeficientes são iguais a zero, assim, a estatística G tem distribuição Qui-Quadrado com p graus de liberdade. Nesse caso L_c é a verossimilhança do modelo com as p variáveis explicativas e L_s é a verossimilhança do modelo apenas com o intercepto.

Teste de Wald

Vamos considerar a seguinte hipótese:

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases}$$

Para testar a hipótese acima, a estatística de Wald é obtida da seguinte forma:

$$W_j = \frac{\hat{\beta}_j}{\widehat{DP}(\hat{\beta}_j)}.$$

onde o desvio padrão é obtido como acima. Se não rejeitarmos H_0 , temos que a variável X_j não explica a variável resposta.

De forma equivalente, teste Wald também pode ser obtido pela multiplicação dos seguintes vetores:

$$W = \widehat{\beta}'[\widehat{\Gamma}(\widehat{\beta})]^{-1}\widehat{\beta} = \widehat{\beta}'(X'VX)\widehat{\beta}.$$

Em que $\widehat{I}(\beta)$ é a matriz da informação de Fisher estimada, apresentada na estimativa do desvio padrão do modelo de regressão logística simples.

Intervalos de Confiança na Regressão Logística Múltipla Discutimos intervalos de confiança para os parâmetros, logito, valores ajustados e odds ratio no modelo de regressão logística simples. Os métodos utilizados no caso de apenas uma covariável são os mesmos no caso da regressão múltipla.

Intervalo de Confiança para os parâmetros

O intervalo de confiança para um parâmetro β_j é baseado em seu respectivo teste de Wald. O intervalo de confiança de $100(1 - \alpha)\%$ para o parâmetro β_j é:

$$IC(\beta_j, 1 - \alpha) = [\widehat{\beta}_j - z_{1-\alpha/2}DP[(\widehat{\beta}_j)]; \quad \widehat{\beta}_j + z_{1-\alpha/2}DP[(\widehat{\beta}_j)]].$$